

$||(\text{mouse})$

IMAGE CONSORTIUM (threelimage.llnl.gov) for further information.
MGI:431042

1. 400
/organism="Mus musculus."

```
/clone_11b="Barstead mouse pooled organs MPLRB4"
```

```

/clone_ilb="barstead mouse pooled organs APUKB4"
/sex="mixed"
/tissue_type="pooled organs"
/dev_stage="7 day"

```

```

/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: EcoRI; Site_2: NotI; lst

```

TGTTACGAATCTGAAGTGGGAGCGCCGCTTTTTTTTTTTTTTTTTTTTTT
3'); double-stranded cDNA was ligated to Eco RI adaptors
[GTTCGATTGGTACC], digested with Not I and cloned into

Library constructed by Bob Barstead.
140 a 126 c 120 g 82 t

cores: 687.00 Length: 124

Ratio:	5.540	Gaps:	0
Similarity:	100.000	Percent Identity:	97.581

50D-12 x AA253897 ..
/1 to: AA253897 from: 1 to: 468

AlaAsnValHisGlnGluAsnGluGluMetGluGlnProLeuGlnAs 17
GCCAAATGTCCACCAGGAAACGAAGAGATGGAGAGCCCCCTGCAGAA 98

YGLuGluAspArgProValGlyGlyGlyGluGlyHisGlnProAlaG 34
ACAAATAAGACCGCCCTGTGGAGAGAGGTGAGGCCACCAGCTGCTG 148

SNASNAASNAASNAASNAHISASNAHISASNAHISARGARG 50
|||||
ACAAACAACAACAACCAACCAATACCAACCAACCAACCAAGA 198

GlnAlaArgArgLeuAlaProAsnPheArgTyrAlaIleProAsnAr 67

nmMetAsnAspGlyLeuGlyGlyAspGlyAspAspMetGluMetPheM 84

luGluMetArgGluIleArgArgGlysLeuArgGluLeuGlnLeuArg 100

CysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisAspHisH1 117

pGluPheCysLeuMetPro 124
TGTCTACGGCTCCTATATGGGGAGCTGTCTAACCAACCACCGATCACCA 398

Tue Mar 12 09:01:17 2002 102(b)
11(morse)

US-09-327-750D-12 x BG088461

Allign seg 1/1 to: BC088461 from: 1 to: 599

1 MetAAsnValHisGlnGluAsnGluGluMetGluInProLeuGlnAs 17
175 ATGGCCAAATCTCCACAGGAACAAGAGAGATTGGAGCAGCCCTCGAGA 224
17 nGlyGluAspArgProValGlyGlyGlyGlyGlyGlyGlyGlyGly 34
225 TGAGCAAGAGACGCCCTGTGGAGAGGTGGAGGTCACCACCTCTG 274
34 TYGASNAAGSADNHSANHISANNHISANNHISANNHISANNHIS 50
275 CAAAC 324
51 GYGlnAlaArgLeuAlaProAsnPheArgTTPAlaIleProAsnAr 67
325 GCCCAGGCCTGCCGACTTCCCTTAACCTCCGATGGGCGCATCCCAAC 374
67 gClnMetAspGlyLeuGlyGlyGlyGlyGlyGlyGlyGlyGly 84
375 GCAGATGANTACGGCTTGCTGGATGGATGGATGGATGGATGGAT 424
84 eTGluGluMetArgGluIleArgGlyLeuArgGlyGluGluGlu 100
425 TGGAGAGAGTAGAGAGATCGCGGAGAAGCTTAGGGAGCTACAGCTGA 474
101 AAnCySLeuArgIleLeuMetGlyGluLeuSerAsnHisHisPhisH 117
475 ANTTGCTCACCTTCTATGGCGAGCTGTCTAACCCACACGATCACCA 524
117 sAspGluPheCysLeuMetPro 124
525 TGATGANTCTCGCTTATGCTCT 546

seg_name: gb_estl: AA272375

seq_documentation_block: 612 bp mRNA EST
AA272375 vbc2207.r1 Barsted mouse pooled organs NPLBB4 26-MAR-1993
Clone IMAGE:761628 5' similar to gb:M8188 OVARIAN GRANULOSA CELL
AA272375 PROTEIN HGR74 (HUMAN);, mRNA sequence.

ACCESSION AA272375.1 GI:1910706
VERSION 1
KEYWORDS EST.
SOURCE house mouse,
ORGANISM Mus-musculus
REFERENCE Enkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;
L (bases 1 to 612)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubouq,T.,
Gelsel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Thompson,K., Stepcock,R., Tan,F., Underwood,K., Moore,B.,
Waterson,W., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mousetest@wustl.edu
This clone is available royalty-free through LLNL ; contact the
LLNL Information Portorium (infoimage.llnl.gov) for further information.
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 507.
Location/Qualifiers
1 . 612 /organism= "Mus musculus"

FEATURES
SOURCE

11 (mouse)

```
101 AnCysLeuAtgTtLeuMetGlyGluLeuSerAnHsHsHsAspHsHsHs 117
|||||
478 AATTGCTTACCGATCTTATGGGGAGCTGCTTACCAACCCAGGATCAACA 527
|||||
117 sAspGluPheCysLeuMetPro 124
|||||
528 TGATGAATTCGCTTATGCTT 549
seq_name: gb_est1:AI152323

seq_documentation_block:
LOCUS AI152323 430 bp mRNA EST 30-SEP-1998
UD9802.r1 Soares_mammary_gland_NMLMG Mus musculus mammary_gland
IMAGE:1477034.5 similar to gb:M8188 OVARIAN GRANULOSA CELL 13.0
KD PROTEIN MGR74 (HUMAN); mRNA sequence.
AI152323
ACCESSION AI152323.1 GI:3680792
VERSION EST
KEYWORDS
SOURCE
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 430)
AUTHORS Marra, M.; Hillier, L.; Allen, M.; Boules, M.; Dietrich, N.; Dubuque, T.;
Geisel, S.; Kucaba, T.; Lacy, M.; Le, M.; Martin, J.; Norris, M.;
Schellenberg, K.; Steptoe, M.; Tan, P.; Underwood, K.; Moore, B.;
Thelings, B.; Wylie, T.; Lennon, G.; Soares, G.; Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
4441 Picot
University School of MedicineP
4441 Picot
University Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:925390
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 419.
Location/Qualifiers
1..430
/organism="Mus musculus"
/clone="IMAGE:1477034"
/clone.lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
pT73D-Pac vector. Library is normalized.
Library was
mediated and purified by Benzo, Soares and M. Fatima Bonaldo."
BASE COUNT 135 a 111 c 113 g 71 t
ORIGIN

alignment_scores:
Quality: 681.00 Length: 124
Ratio: 5.537 Gaps: 0
Percent similarity: 99.194 Percent identity: 97.581

alignment_block:
US-09-327-7500-12 x AI152323
Align seg 1/1 to: AI152323 from: 1 to: 430
```

```
1 MetAlaSerValHisGlnGlnGlnGluGluMetGluGlnProLeuGlnAs 17
|||||
54 ATGGCCATGTCTCCAGGAGGAGGATGGAGCCCTGCACGA 103
|||||
17 nGlyGluGlnAspArgProValGlyGlyGlyGlnHisGlnProAlaG 34
|||||
104 TGGACAGGAGACCCGCTCTGTGATGGAGGTGAGGGCCACCAGCTGCTG 153
|||||
34 lyAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisAsnHisArg 50
|||||
154 CAACACACACACACACACACACACACACACACACACACACACACAC 203
|||||
51 GlyGlnAlaArgLeuAlaProAsnProAsnProAsnProAsnProAsn 67
|||||
204 GCCCAGGCTGCGCGACTTGGCCCTTACTTCGATGGCCATTCGCACAG 253
|||||
67 gGlnMetAsnAspGlyLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyG 84
|||||
254 GCAGATGAATGACGGGTGGGTGGAGATGAGATGATGGAATGTCTCA 303
|||||
84 eTgGluMetArgGluLeuArgArgGlyLeuArgGlyLeuArgGlyLeu 100
|||||
304 TGGAGGAGATGACAGAGATCCGAGAGAGCTTAGGGAGCTACAGCTGAGA 353
|||||
101 AsnGlyLeuArgGlyLeuMetGlyGlyGlyGlyGlyGlyGlyGlyGly 117
|||||
354 AATTGCTACGATCTTATGGGGAGCTGCTTACCAACCCAGGATCAACA 403
|||||
117 sAspGluPheCysLeuMetPro 124
|||||
404 TGATGAATTCGCTTATGCTTATGCTTATGCTTATGCTTATGCTTAT 425
seq_name: gb_est1:AM908751

seq_documentation_block:
LOCUS AM908751 503 bp mRNA EST 25-MAY-2000
UD9802.r1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
IMAGE:1477034.5 similar to gb:M8188 OVARIAN GRANULOSA CELL 13.0
KD PROTEIN MGR74 (HUMAN); mRNA sequence.
AM908751
ACCESSION AM908751.1 GI:8073984
VERSION EST
KEYWORDS
SOURCE
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 503)
REFERENCE NC1-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Unpublished (1997)
JOURNAL
COMMENT Other ESTs: u57405.x1
Contact: Robert Strausberg, Ph.D.
Email: cygphs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:942292
Seq primer: -408P from Gibco
High quality sequence stop: 468.
Location/Qualifiers
1..503
/organism="Mus musculus"
/clone="IMAGE:1477034"
/clone.lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Eco RI sites of the
pT73D-Pac vector. Library is normalized.
Library was
mediated and purified by Benzo, Soares and M. Fatima Bonaldo."
BASE COUNT 135 a 111 c 113 g 71 t
ORIGIN
```

2

22, 23

[illegible]

```

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/db_xref="dbEST:AA272375"
/chromosome="x"
/map="near Pip"
/tissue_type="pooled organs"
1..854
/gene="Bex3"
172..546
/gene="Bex3"
/feature="Bex3-1"
/product="brain expressed X-linked protein 3"
/protein_id="AA024431.1"
/db_xref="GI:4580594"
/translation="MANVHQENEMDPLQGGEDRPVGGGEGGQPAANNNNNNHN
HNHFRGQARLAPENFMAIRQNMNDLGGQDDMMEMEMREIRKKLRLEIQLRNC
LRTLGLGSHHHHHHDFCLAP"
237 a 212 c 228 g 177 t
BASE COUNT

```

ORIGIN

[illegible]

```

seq_name: gb_ro:AF187065

seq_documentation_block:
LOCUS       AF187065             519 bp          mRNA
DEFINITION  Rattus norvegicus p75NTR-associated cell death executor. (Nade)
            mRNA complete cds.
ACCESSION   AF187065
VERSION     AF187065.1  GI:8452895
KEYWORDS    .
SOURCE      Norway rat.
             Rattus norvegicus
             Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Pungidae 1 to 519)
REFERENCE   1 (bases 1 to 519)
AUTHORS    Suwanto, P, Hanokita, T, Shoji, Hoshino, S., Kimura, M.T., Nadano, D.,
            Watanabe, T, Irie, S., Greene, I.A. and Sato, T.A.
TITLE      NADE, a p75NTR-associated cell death executor, is involved in
            signal transduction mediated by the common neurotrophin receptor
            p75NTR
JOURNAL    J. Biol. Chem. 275 (23), 17566-17570 (2000)
MEDLINE    20298829
REFERENCE   2 (bases 1 to 519)
AUTHORS    Mukai, J., Hachiyu, T., Hoshino, S., Kimura, M., Nadano, D., Suwanto, P.,
            Hanokita, T., Li, I., Irie, S. and Sato, T.
TITLE      Directed p75NTR-SP-1999 Ocularyngology/Pathology, Columbia
JOURNAL    University, 630 West 168th St., P&S 11-451, New York, NY 10032, USA
            Location/Qualifiers
FEATURES             1..519
             source

```

```

US-09-327-730D-12 x BG08B461
Align seg 1/1 to: BG08B461 from: 1 to: 599

1  MetAlaenValHisGlnGlnAsnGluUmetGluGlnProLeuGlnAs 17
175  ATGCCCATATGTCACCAAGAAAGAAAGAGATGAGACAGCCCTCGAGA 224
17  GClGylGluAspArgProValGlyGlyGlyGlyGlyGlyGlnProAlaG 34
225  TGGACAGAGAGACCCCTCTGTGGAGAGAGATGAGGCCACCACTCTGCTG 274
34  lYasnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisHisAsArgSp 50
275  CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 324
51  GylGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnAr 67
325  GGCACAGCTCGCCGACATCTGCCCTTAACCTCCGATGGGGCATTCGCCAC 374
67  GClMetAsnAspGlyLeuGlyGlyGlyGlyAspGlyAspMetGluMetPhe 84
375  GCACAGTAAATGACGGCTGTGGTGGAGATGGAGATGATATGGAATGTTC 424
84  eGluGluMetArgGluIleArgArgLysLeuArgGlyGluGluGluLeuArg 100
425  TGCAGCAGATGAGAGAGATCCGGGAGAACCTTGGGAGCTACAGCTGAGA 474
101  AenCysLeuArgTlleLeuMetGlyGlyLeuSerAsnHisHisAspHisH 117
475  AATGTGCTACGACATCTTATGCGGGAGCTCTCAACCAACCAACGATCA 524
117  sAspGluPheCysLeuMetPro 124
525  TATGATGATTCCTCTTATGCTT 546

seq_name: gb_est1:AA272375

seq_documentation_block:
LOCUS       AA272375             612 bp             mRNA
DEFINITION  vb62077.r1 Barated mouse pooled ocsang MFLRB4 Mus musculus CDNA
            clone IMAGE:761628 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
            13.0 KD PROTEIN HGR74 (HUMAN); mRNA sequence.
ACCESSION   AA272375
VERSION     AA272375
RECORDS     1
FEATURES             (1) GI:1910706
ORIGIN
1  ATGTTGCTACGACATCTTATGCGGGAGCTCTCAACCAACCAACGATCA
2  CCAACCAACCAACGATCAACCAACCAACCAACCAACCAACCAACCAAC
3  CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA
4  ACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
5  CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
6  CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
7  CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
8  CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
9  CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
10 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
11 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
12 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
13 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
14 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
15 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
16 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
17 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
18 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
19 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
20 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
21 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
22 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
23 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
24 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
25 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
26 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
27 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
28 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
29 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
30 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
31 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
32 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
33 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
34 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
35 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
36 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
37 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
38 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
39 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
40 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
41 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
42 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
43 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
44 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
45 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
46 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
47 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
48 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
49 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
50 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
51 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
52 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
53 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
54 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
55 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
56 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
57 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
58 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
59 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
60 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
61 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
62 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
63 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
64 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
65 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
66 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
67 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
68 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
69 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
70 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
71 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
72 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
73 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
74 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
75 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
76 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
77 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
78 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
79 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
80 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
81 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
82 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
83 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
84 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
85 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
86 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
87 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
88 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
89 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
90 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
91 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
92 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
93 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
94 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
95 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
96 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
97 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
98 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
99 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
100 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
101 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
102 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
103 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
104 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
105 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
106 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
107 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
108 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
109 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
110 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
111 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
112 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
113 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
114 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
115 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
116 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
117 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
118 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
119 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
120 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
121 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
122 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
123 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
124 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
125 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
126 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
127 CCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC
128 CCAACCAACCAACCA
```


[illegible]

```

seq_documentation_block:
:
: Sequence 1 Application PC/TUS9206840
:
: GENERAL INFORMATION:
:
: APPLICANT: Shi, Yang
: APPLICANT: Seto, Edward
: APPLICANT: Shenk, Thomas
:
: TITLE OF INVENTION: Y1 TRANSCRIPTION FACTOR AND METHODS OF
:
: TITLE OF INVENTION: ISOLATING SAME
:
: NUMBER OF SEQUENCES: 10
:
: CORRESPONDENCE TO:
: ADDRESSEE: Oatroltek, Paper, Gerb & Soffen
: STREET: 1180 Avenue of the Americas - 77th Floor
:

```

CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10046-8403
 COMPANY: COMPTON AMERICA CORP.
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US92/06840

APPLICATION NUMBER: PCT/US92/06840
 FILING DATE: 19920814
 CLASSIFICATION: C12N 2501/00
 CUSTODIAN: AU 1805
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/746,485
 FILING DATE: 16-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Dennis, Manette
 REGISTRATION NUMBER: 30,623
 REFERENCE/DOCKET NUMBER: M-12594 CIP (1370-8)
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (214) 382-0700
 TELEFAX: (214) 382-0888
 TELEX: 236525
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2353 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 RECOMBINATION: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 TISSUE TYPE: HeLa cells derived from cervical
 TISSUE TYPE: carcinoma
 CELL TYPE: tumor cells
 CELL LINE: HeLa
 IMMEDIATE SOURCE:
 LIBRARY: D96/AH-2
 CLONE: p4-1 or pY1
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 241..1485
 PCT-US92-06840-1

alignment_scores:
 Quality: 84.50
 Ratio: 1.7951
 Percent Sim: 99.03
 Gaps: 00.00
 Percent Var: 00.00

alignment_scores:		
Quality:	84.50	Length: 86
Ratio:	1.798	Gaps: 3
Percent Similarity:	54.55	Percent Identity: 30.07

[illegible]

```

; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 663..3164
; US-08-396-001-3

alignment_scores:
  Quality: 79.00      Length: 92
  Ratio: 1.362      Gaps: 5
  Percent Similarity: 63.043      Percent Identity: 26.087

alignment_block:
  US-09-327-750D-12 x US-08-396-001-3 ..
  Align seg 1/1 to: US-08-396-001-3 from: 1 to: 3455

17 Asnglyglucluaspaqprvaiglyglygluyluaglnproal 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2451 AACACGAAATAGAACCC...CATACAAAAA 2482
33 aglyanasnasnasnasnasnasnasnasnasnasnasnasn 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2483 TAGTCATATCATATCATATCATATCATATCATATCATATCA 2532
50 rfglyglnlaargargleuAlaProasnPheargTrpAlailePro... 65
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2533 ATACATATATATCAAGAGTATACCCGTCATTTTCTTACCA 2582
66 .....AsnArgGlnMetAsnAspGlyLeuGlyAspGlyAsp 79
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2583 ATGCTTACCATAGAGAGTACAGCTCTGTACCAATAT..... 2624
79 pMetGluMetPheMetGluGluMetArgGluileargArglysLeuarg 95
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2625 .....TTCTCAACCAATATG.....CACAGATCAGA 2652
96 GluLeuGlnLeuArgAsnCysLeu 103
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2653 AAATTCATCTCTCCCAACCAATTA 2676

seq_name: /cgn2_6/pdata/2/lna/6A_COMB.seq:US-08-155-888-1

seq_documentation_block:
; Sequence 1, Application US/08155888
; Patent No. 6218512
; GENERAL INFORMATION:
; APPLICANT: Hoffmann, Stephen L.
; APPLICANT: Hedstrom, Richard C.
; APPLICANT: Sedegah, Martha
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE PROTECTIVE
; AGAINST MALARIA, METHODS OF PROTECTION AND VECTOR FOR
; DELIVERING POLYNUCLEOTIDE VACCINES
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Naval Medical Res. & Dev. Cmd.
; STREET: Bldg. 1, T-12 8901 Wisconsin Ave.
; CITY: Bethesda
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20889-5606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/155,888
; FILING DATE:
; PRIORITY DATE:
; PRIORITY NUMBER: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Speveck, A. David
; REGISTRATION NUMBER: 24,743
; REFERENCE/DOCKET NUMBER: N.C. 75,851
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 295-6759

```


XX WPI: 2000-611515/38.
 DR P-PSDB: AAB58845.
 XX New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases.
 XX
 PS Claim 1: Page 581-582; 1295pp; English.
 XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterization of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytotoxic; immunosuppressive;
 CC neoplastic; neuroprotective; antiviral; antileukic; hepatotropic;
 CC antidiabetic; antifungal; antiparasitic and cardiant activity. The
 CC antibodies, particularly breast and ovarian cancer. The nucleic acid sequences,
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and antagonists may also be used in the diagnosis,
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune hemolytic anemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemias; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.
 XX
 SQ Sequence 917 BP; 228 A; 203 C; 270 G; 211 T; 5 other;

alignment_scores:
 Quality: 550.50 Length: 119
 Ratio: 5.193 Gaps: 1
 Percent Similarity: 89.076 Percent Identity: 84.874

alignment_block:

US-09-327-750D-12 x AAF21748 ..

Align seg 1/1 to: AAF21748 from: 1 to: 917

6 GlnGluAsnGluGluMetGluGluProLeuGluAsnGluGluAspArg 22
 306 CAGGAAACGAGAGATGGAGCAGCTATGCAAGATGGAGAGGAGACCG 355
 22 gProValGlyGlyGlyGluGluGluGluGluGluGluGluGluGluGluGlu 39
 356 CCCTTTGGAGAGGAGTGAAGCCACAGCCGCTGCAGAAAT..... 395
 39 anAsnHisAsnHisAsnHisAsnHisAsnHisAsnHisAsnHisAsnHis 55
 396CGACGGGAGAGCTGCGCGA 416
 56 LeuAlaProAsnPheArgTTPAlaLeuProAsnArgGlnMetAsnAspG 72
 417 CTTCGCCCTATTTTCGATGGCCATACCCATAGGAGATCAATGATGG 466
 72 yLeuGlyGlyAspGlyAspAspMetGluMetPheMetGluGluMetArg 89
 467 GATGGGTGGAGATGGAGATGATGGAATATTCATGGAGAGATGAGAG 516
 89 luLeuArgArgGlyLeuArgGluLeuGluGluGluGluGluGluGluGlu 105
 517 AATCAGAGAAACTTAGGAGCTCCGCTTGAGGAAATGTCTGCTATC 566
 106 LeuMetGlyGluLeuSerAsnHisHisAspHisHisAspGluPheCysLe 122
 567 CTATGGGGGGAGCTCTCTATCACCATGACCATCATGATGAATTTTGCCT 616
 122 uMetPro 124

|||||
 617 TATGCT 623
 seq_name: /SIDS2/gcdata/geneseq/geneeqn/NA2000.DAT:AAC01005
 seq_documentation_block:
 XD AAC01005 standard: cDNA; 532 BP.
 XX
 AC AAC01005;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST, SEQ ID NO: 1003.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; as.
 XX
 XS Homo sapiens.
 XX
 PN XP1033401-A2.
 PD 06-SEP-2000.
 PF 21-FEB-2000; 2000EP-0200610.
 PR 26-FEB-1999; 99US-0122487.
 XX (GEST) GENSET.
 PA
 XX Dumas Milne Edwards J, Duclert A, Giordano J;
 XX WPI: 2000-500381/45.
 DR P-PSDB: AAG00999.
 XX

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 1: SEQ ID 1003; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence that encodes a protein of 1003 amino acids. The mRNA
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC cDNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 XX vectors.

SQ Sequence 532 BP; 151 A; 121 C; 163 G; 91 T; 6 other;

alignment_scores:
 Quality: 504.50 Length: 117
 Ratio: 5.096 Gaps: 1
 Percent Similarity: 84.615 Percent Identity: 79.487

alignment_block:

US-09-327-750D-12 x AAC01005 ..

Align seg 1/1 to: AAC01005 from: 1 to: 532

1 MetAlaAsnValHisGlnGluAsnGluGluMetGluGluProLeuGlnAs 17
 219 ATGGCAATATTATCCAGGAAACCAAGAGATGGAGCGCTATGCAGAA 268
 17 nGlyGluGluAspArgProValGlyGlyGlyGlyGluGlyHisGlnProAlaG 34

553 TATATGCGCTTCCAAACTCTGACCTGACACCACTTATGAC..... 594

118 pGluPheCysLeuMetPro 124

|||||.....

595TTTTGCTCATACCT 609

seq_name: /SIDS2/jcgdata/geneseq/geneseq/NA2000.DAT:AAH87147

seq_documentation_block:

ID AAH87147 standard: DNA: 187 BP.

XX AAH87147:

XX 08-JAN-2001 (first entry)

DE Ret hepatocyte carcinogenesis biomarker nucleic acid SEQ ID NO:71.

KW Rat: phenobarbital: carcinogenesis marker: carcinogenesis: detection;

KW identification: carcinogenic: probe: primer: ds.

OS Rattus norvegicus.

XX WO200044902-A2.

XX 03-AUG-2000.

XX 28-JAN-2000; 2000WO-US00503.

XX 29-JAN-1999; 99US-0118078.

XX (SEAR) SEARLE & CO G D.

XX Bunch RT, Curtis SW, Rodi CP, Morris DL;

XX WPI: 2000-505977/45.

XX New nucleic acid encoding a carcinogenic biomarker, induced by

PT phenobarbital treatment of rat hepatocytes, useful for identifying

PT carcinogenic compounds.

PS Claim 1; Page 73; 240pp: English.

CC AAH87080 to AAH87656 represent nucleic acid sequences (N1) encoding a

CC carcinogenesis biomarkers. The carcinogenesis biomarkers are induced by

CC treating rat hepatocytes with phenobarbital. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acids are

seq_name: /SIDS2/jcgdata/geneseq/geneseq/NA2001.DAT:AAH45143

seq_documentation_block:

ID AAH45143 standard: cDNA: 792 BP.

XX AAH45143:

XX 07-SEP-2001 (first entry)

XX Human brain expressed X-linked protein, hbex, coding sequence.

KW Human: brain expressed X-linked protein; cytostatic; auditory; neutrophic;

KW hbex; dysmrioplasia; hereditary disease; cancer; tumour; deafness;

KW X-chromosome-binding mental retardation; lissencephalous disease; ss.

OS Homo sapiens.

XX WO200140286-A1.

XX 07-JUN-2001.

XX 27-NOV-2000; 2000WO-CN00502.

XX 30-NOV-1999; 99CN-0124179.

XX (BIOR-) BIORAD GENE DEV LTD SHANGHAI.

XX Mao Y, Xie Y;

XX WPI: 2001-397944/42.

XX P-PSDB: AAB9224.

XX Isolated human brain-expressed X-linked polypeptide used to diagnose

XX and treat of dysmrioplasia, hereditary diseases, cancer, tumor,

XX deafness and X-chromosome-binding mental retardation.

PS Claim 5; Page 22; 30pp: Chinese.

XX The present sequence is the coding sequence for a human brain-expressed

XX X-linked protein (hbex). hbex and its coding sequence are useful in the

XX diagnosis and treatment of dysmrioplasia, hereditary diseases, cancer,

XX tumours, deafness, X-chromosome-binding mental retardation, autism,

XX lissencephalous disease, hbex. As a nucleic acid, the present sequence

XX can be used for the diagnosis and treatment of dysmrioplasia, hereditary

XX diseases, cancer, tumor, deafness, X-chromosome-binding mental

XX retardation, or the treatment of dysmrioplasia, hereditary diseases,

XX cancer, tumor, deafness, X-chromosome-binding mental retardation,

XX autism, lissencephalous disease, hbex. As a nucleic acid, the present

XX sequence can be used as primers or probes, or in producing

XX gene chips or microarrays.

XX Sequence 792 BP; 214 A; 172 C; 219 G; 187 T; 0 other;

alignment_scores:

Quality: 184.00 Length: 131

Ratio: 2.435 Gaps: 5

Percent Similarity: 57.252 Percent Identity: 36.647

alignment_block:

US-09-327-750D-12 x AAH45143 ..

Align seg 1/1 to: AAH45143 from: 1 to: 792

1 MetAlaAsnValHisGlnGluAsnGluGluMetGluGlnProLeuGlnAs 17

||| |||||.....

212 ATGGAAATGCGCAACCAAGAAATGAGAAAGGAGCAGTCTCTAATPA 261

17 nGlyGlu...AspArgProValGlyGlyGlyGlyHisGlnProk 33

||| |||||.....

262 AGGGAGCCCTTGGCCCTCTTGGATGCTGGTACTGTCTGCTTA 311

33 LacIyAnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisAsnHisAsn 49

||| |||||.....

312 GAGGAAT.....CGTAGG 325

XX	AAFS9611	standard; CDNA; 898 BP.
XX	AAFS9611;	
XX	24-APR-2001	(first entry)
XX	Human	cell cycle and proliferation protein CCYPR-22 cDNA, SEQ ID NO:76.
XX	Cell cycle and proliferation protein; CCYPR; human; agonist;	
XX	antagonist; gene therapy; detection; gene therapy;	
XX	developmental disorder; cell signalling disorder;	
XX	cell proliferative disorder; cancer; tumour; anaemia;	
XX	arteriosclerosis; asthma; allergy; diabetes mellitus;	
XX	menstrual cycle disorder; bacterial infection; ss.	
XX	Homo sapiens.	
XX	OS	
XX	PN	WO200107471-A2.
XX	XX	01-FEB-2001.
XX	XX	
XX	PF	21-JUL-2000; 2000WO-US19948.
XX	XX	

PK 21-JUL-1999; 99US-0153073.
PK 08-SEP-1999; 99US-0153129.

XX	..(INCY-) INCYTE GENOMICS INC.	
PA		
XX		
XX	Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;	
PI	Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;	

XX
DR
WPI: 2001-112727/12.

Human cell cycle and proliferation proteins and polynucleotides are

PT signaling disorders and cell proliferative disorders including cancer -

Sequences AF595900-AF59643 represent cDNAs encoding 54 human cell cycle and proliferation proteins (CCVPR). AA860453-AA860506. CCVPR and agonists of CCVPR are used to treat diseases or conditions associated with decreased expression of functional CCVPR, while CCVPR antagonists are used to treat diseases or conditions associated with overexpression of functional CCVPR. Monoclonal or polyclonal antibodies to CCVPR may be used in enzyme-linked immunosorbent assays (ELISA) or

CC recombinant antibodies, oligonucleotides and proteins (receptors)
CC that specifically bind to CCypr and in drug screening methods to
CC identify compounds that modulate the activity of CCypr. CCypr
CC nucleotides can be used to generate transgenic animal models of human
CC disease, and can be used in gene therapy in target cells with genetic
CC abnormalities with respect to the expression of CCypr. For the
CC treatment or prevention of a disorder associated with CCypr,
CC diseases which can be diagnosed, treated and prevented using CCypr

CC proteins, nitric acid, agonists of antagomists and dopamine,
CC developmental and cell signalling disorders, and cell proliferative
CC disorders including cancer. Specific examples of these disorders
CC include anaplasia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
CC diabetes mellitus, disorders of the menstrual cycle and infections
XX caused by bacteria.
XX
SQ Sequence 898 BP; 250 A; 186 C; 251 G; 211 T; 0 other;

alignment scores:

Ratio: 2.287 Gaps: 6

PERCENT SIMILARITY: 30.010 PERCENT IDENTITY: 30.304

alignment_scores:		
Quality:	171.50	Length: 132
Ratio:	2.287	Gaps: 6
Percent Similarity:	56.818	Percent Identity: 36.364

obtained the full 5' UTR is rarely included, 5' ESTs are derived from mRNA with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain upstream regulatory sequences and to design expression and secretion vectors.

```

alignment_scores:      Length: 126
                      Quality: 162.50
                      Ratio: 2.355
                      Gaps: 4
Percent Similarity: 54.762 Percent Identity: 35.714

alignment_block:
US-09-327-750D-12 x AAC03880 ..

Align seq 1/1 to: AAC03880 from: 1 to: 662

3 AsnValHisGlnGluAn.....GluGluMetGluGlnProLeuG1 16
252 ANTGAGTCACAGGAAATGCTGTAAGAAAGCTGAAAGAGGACCTACAGCTACCTAA 301
16 nAsnGlyGluGlu...AspATgProValGlyGlyGlyGluHisGlnP 32
302 TAAAGGGGAGCCCTTGCCCTACCTTTTGATGTGTGTGTAAGTACTGTGTC 351
32 rGAlGgLYAsnAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisH18 48
352 CTAGAGGAAT.....CGT 365
49 ArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTTPAlaLepr 65
366 AGGCGTTTCGGGTGTGGCCAGCCACCCATCTCTCATGTATGAGTGGATATGT 415
65 oAsnArgGlnMcAsnAspGlyLeuGlyGlyAspGlyAspAspMetGluM 82
416 GCATAGGCTTGGAGACCAAGCAGGACAGTGANAGAGAGATATGGA 465
82 ePheMetGlyGluMetArgGluGluLeuArgArgGlyLeuArgGluLeuGln 98
466 GGATTTGGGAGAGAGGTGTGACACACTGTATGGAAGAAAGCTGAGGGAAGACAG 515
99 LeuArgAsnCysLeuArgTLeuLeuMetGlyGlyLeuLeuSerAsnHisHisAs 115
516 TTGATGCTATAGTCTGGCGGCGAGTGCACCTACG.....CCCTCTCACCTGCA 562
115 pHHisLeuArgGluPheCysLeuMetPro 124
563 CCATCATGATGATGATTTTGCNNNNATGCC 590

```

ZZ/ TGGACAGGAAAGACCGCCCTGGTGGGAC

```

ORGANISM      Homo sapiens
REFERENCE      Bkaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
AUTHORS       Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE         1 (bases 1 to 793)
JOURNAL       Direct submission
REMARK        Submitted (13-FEB-2001) National Institutes of Health, Mammalian
COMMENT       Gene Collection (MGI), Cancer Genomics Office, National Cancer
              Institute, 31 Center Drive, Room 11A33, Bethesda, MD 20892-2950,
              USA.
              NIH-MGC Project URL: http://mgc.nci.nih.gov
              Contact: MGC help desk
              Email: gcapsb-remail.nih.gov
              Tissue procurement: ATCC
              cDNA Library Preparation: Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLML)
              DNA Sequencing by: Genome Sequence Centre,
              BC Cancer Agency, Vancouver, BC, Canada
              Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
              Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
              Letticia Hsiao, Martin Krzyzinski, Reta Kutsche, Oliver Lee, Soo
              Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
              Nens, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Seedi, Jacqueline
              Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Scott,
              Michael Thorne, Natasha Tsai, Natasha Vasya, Scott Zuyderdyk,
              George Yang, Scott Zuyderdyk, Marco Marra.
              Clone distribution: MGC clone distribution information can be found
              through the I.M.A.G.E. Consortium/LLML at: http://image.llnl.gov
              Series: IPAL plate; 6 Row; k Column: 22
              This clone was selected for full length sequencing because it
              passed the following selection criteria: matched mRNA gi: 7657043.
              Location/Qualifiers
                  1..793
                  /organism="Homo sapiens"
                  /db_xref="LocustID:27018"
                  /db_xref="taxon:9606"
                  /cds_start="1"
                  /cds_end="793"
                  /cds_type="cDNA"
                  /clone_lib="NIH-MGC_16"
                  /lab_host="DH10B-R"
                  /note="Vector: pOTB7"
                  196..531
                  /cdon_start=1
                  /product="p53MR-associated cell death executor; ovarian
                  granulosa cell protein (p130)"
                  /protein="p130"
                  /protein_id="P130231"
                  /translation="MANIHNENMEPMQNGEEDRPLGGSGHDPAGNRGQARLA
                  /PFRNATPQINDMGGGDDMEIFMEENREIRKLRELQRLRLMGSLHSD
                  HIDEFLAP"
BASE COUNT    225 a 170 c 205 g 193 t
ORIGIN
alignment_scores
    Quality: 576.50
    Length: 124
    Identity: 5194
    Gaps: 1
    Percent Similarity: 89.516
    Percent Identity: 84.677
alignment_block
    US-09-327-7500-12 x BC003190
Align seg 1/1 to: BC003190 from: 1 to: 793
1 MetAaenValHigInluagngluGluwecGluInProLeuGlnas 17
196 ATGCACATATTCACAGGACACAGAGATAGATGAGCGCTATGCAGAA 245
17 nGlyVluGlnAspGProValGlyGlyGlyGluYHsGlnProAlag 34
246 TCGAGAGACAGACGCCCTTTTGGGAGGAGGTGAAGCCACGACCTCGAC 295

```

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..891
/gene="Nape"
312..640g
/contig="640g"
/notes="Involved in the common neurotrophin receptor
p75NTR-mediated signal transduction."
/codon_start=1
/product="p75NTR-associated cell death executor"
/protein_id="AAF5129.1"
/db_xref="GI:8452894"
/translation="MANHSENEEMQPMQNEEDRPLGGEGHOPAGNRQOARLA
PNFRRAIPNQINDGGDDDDHETFEHREIRKRLHQLANCLRLMGLSELNHD
HIDEFLMP"
BASE COUNT 251 a 182 c 224 g 234 t
ORIGIN

alignment_scores:
  Quality: 576.50      Length: 124
  Ratio: 5.194         Gaps: 1
Percent Similarity: 89.516 Percent Identity: 84.677

alignment_block:
US-09-327-750D-12 x AF187064 ..
Align seg 1/1 to: AF187064 from: 1 to: 891

1 MetAlaAanValHisGluGluAargGluMetGluGluProLeuGlnAs 17
312 ATGGCAATATTCCACGAAAGAGAGAGATGGAGCGCTATGCGAA 361
17 nCylGluAAspArgProValGlyGlyGlyGlyHisGlnProAlaG 34
362 TGGAGGAAGACCGCCCTTTGGGAGGAGGTGAAGCCACCGCTGCAG 411
34 lYAsnAsnAanAanAanHisAsnHisAsnHisAsnHisArgArg 50
412 GAAT .....CCACGG 422
51 GylGlnAlaArgLeuAlaProAnPheArgTrpAlaIleProAnAr 67
423 GGACAGCTCGCGACTGCCCTTATTTTCAGGGCCATACCCATAG 472
67 gClnMetAsnAspGlyLeuGlyGlyVAspGlyAspMetGluMetPhe 84
473 GCAGATCAATGATGGATGGGTGGAGATGAGATGATATGGAATATCA 522
84 eCylGluMetArgGluIleArgGlyLysLeuArgGluLeuGlnLeu 100
523 TGGAGGAGATGAGAAATACAGAGAAACTTAGGAGCTGAGTTGAGG 572
101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisAspHis 117
573 AATTGCTGCTATCTTATGGGGAGCTCTCTATCCATCCATGACCTCA 622
117 sAspGluPheCysLeuMetPro 124
623 TGATGAATTTTGCCTTATGCTCT 644

seq_name: gb_sts:G72708
```

```
seq_documentation_block:
LOCUS G72708 421 bp DNA STS 08-AUG-2001
DEFINITION MARC 4953:4954:99193031:1 SCF - porcine spleen STS
ACCESSION G72708.1
VERSION G72708.1 GI:15146738
KEYWORDS STS.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 421)
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L. and
Keelie,J.W. Nucleotide polymorphism (SNP) discovery in expressed porcine
Single nucleotide polymorphism (SNP) discovery in expressed porcine
Unpublished (2001)
Contact: Freking BA
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4278
Fax: 402 762 4173
Email: freking@mail.marc.usda.gov
Primer A: CCAATGGGTGGAACCTACT
Primer B: CCACGACCAATAGACG
STS Size: 300
PCR Protocol:
Denature: 95 degrees for 15 minutes
Anneal: 95 degrees for 30 seconds
Anneal: 56 degrees
Extension: 68 degrees for 2 minutes
Cycles: 32 to 45
Protocol:
Template: 50-200 ng genomic DNA
Primer: each 20 pmoles
dNTPs: each 88 uM
Taq Polymerase: 0.25 units (Qiagen HotStar)
Buffer: Commercially supplied Qiagen HotStar buffer
The STS is derived from PCR amplicons generated from genomic DNA,
sequenced from each end using the amplification primers. The
sequence does not necessarily represent the entire amplicon.
Sequence derived from PolyPhred was trimmed from each end of each
unique contig until five consecutive bases exceeded a quality score
threshold of 20, and the next 10 bases averaged a quality score of
20 or greater. Amplicon size was estimated by agarose gel
electrophoresis.
Location/Qualifiers
1..421 gb="Sus scrofa"
/stable="white composite, duroc, meishan, minzhu,
fengling, crossbreds"
/db_xref="taxon:9823"
/sex="male and female"
/clone_lib="SCF - porcine spleen"
/dev_stage="adult"
/note="Organ: spleen"
<1..>421
BASE COUNT 170 a 101 c 86 g 102 t 2 others
ORIGIN

alignment_scores:
  Quality: 411.00      Length: 77
  Ratio: 5.338         Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.104

alignment_block:
US-09-327-750D-12 x G72708/rev ..
Align seg 1/1 to reverse of: G72708 from: 1 to: 421

48 HlaArgArgGlyGlnAlaArgLeuAlaProAnPheArgTrpAlaI 64
421 ATATGACGGGACAGCTCGCGACTGCCCTTATTTCCATGCGCCCT 372
64 eProAnArgGlnMetAsnAspGlyLeuGlyGlyVAspGlyAspMetG 81
371 ACCCAATAGCAGATCAATGATGGATGGGTGGAGATGAGATGATATGG 322
```

seq_name: gb_pr:AK000959

seq_documentation_block:

LOCUS AK000959 1229 bp mRNA 22-FEB-2000

DEFINITION Homo sapiens cDNA FLJ10097 fis, clone HEMBA1002458, weakly similar to OVARIAN GRANULOSA CELL 13.0 K D PROTEIN HG974.

ACCESSION AK000959

VERSION GI:7021945

KEYWORDS o1igo capping; fis (full insert sequence).

SOURCE Homo sapiens, 10 weeks whole embryo, mainly head cDNA to mRNA, clone:lib:HEMBA1 clone:HEMBA1002458.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS Nagashima, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugeno, S., Takahashi-Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Tatekuchi, K., Arita, M., Nabekura, T., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagashira, K., Masuho, Y. and Oshima, A.

TITLE NEDO human cDNA sequencing project


```

452 CTGAGGGAAAGACAGTGTGAGTCATATGCTGCGGGCAGTCAGCACTGAC... 499
110 uSerAsnHisHisAspHisHisAspGluPheCysLeuMetPro 124
      |||||||
500 .CCCCCTCACCATGACCATCATGATGAGTGTGCGCTATGCCC 541

```

```

alignment_scores:      Length: 131
                        Quality: 191.00
                        Ratio: 2.513
                        Gaps: 5
Percent Similarity: 58.015 Percent Identity: 37.405

alignment_block:
US-09-327-750D-12 x AF220189    . . .

Align seg 1/1 to: AF220189 from: 1 to: 828

1 MetAlaAsnValHISIGlnAanGluMetGluInProLeuGlnas 17
   ||| |||||:::|||||:::|||||:::|||||:::|||||:::
206 ATGGAAATGCCACCAAGAAATGAAGAAGAAGCAGCATTCCTTAATA 255
   ||| nGlyGlu... AspaArgProValGlyGlyGluGlyHisGlnPro 33
256 AGGGAGACCCTTGCCCTTGGATCGGTGGAATCTGTGTGCCTA 305
   |||||
33 IacIyAanAanAanAanAanAanHISAsnHISAsnHISAsnHISAsn 49
   |||||
306 GAGGAAT.....CGTAGS 319

50 ArgGlyGlnAlaArgAgluAlaProAsnPheArgTrpAlaile.... 64
   ||| ::::::::::::::::::::
320 CGGTTCCGGTTAGCAGCCCATTCTCGAGTAGATAGGATATGCA 369
   |||||

65 .....ProAnArgGlnMetAsnAspGlyLeuGlyGlyAsp 77
   |||||:::|||||:::
370 TAGGCTTGAGAACCCAGCAGGATGAGAA..... 403

77 lyaAspAspMetLueMetPheMetGluGlumetArgGluIleArgArglys 93
   |||||
404 ..GAGAAATGGAAAGGATGGGAGAGGCTGAGACNCTGATGGANAAG 451
   |||||

94 LeuArgGluLeuGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLe 110
   |||||

```


101 seq_name: gb_estl:AA215070
seq_documentation_block: 557 bp mRNA EST 03-FEB-1997
LOCUS AA215070.1 r1 stragene mouse embryonic carcinoma (4937317) Mus
DEFINITION musculus IMAGE:651487 5', similar to gb:M38188 OVARIAN
GRANULOSA CELL 13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
AA215070
ACCESSION AA215070.1 GI:1814831
VERSION
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus.
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
1 (bases 1 to 557)
REFERENCES
Marrin, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and
Waterson, R.
The WashU-HMI Mouse EST Project
Unpublished (1997)
COMMENT
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28ml3 revl. ET from Amersham
High quality sequence stop: 447.
Location/Qualifiers
1. -557
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:651487"
/clone_lib="Stragene mouse embryonic carcinoma (#937317)"
/tissue_type="carcinoma"
/tissue_emb="embryo"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: pBluescript SK-; Site:1: EcoRI; Site:2:
XhoI; Cloned unidirectionally. Primer: Oligo dt. P19 cell
line. Average insert size: 1.0 kb; Uni-ZAP XR vector: -5'
adaptor sequence: 5' GAATTCGGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTT 3'."
BASE COUNT 159 a 145 c 165 g 88 t
ORIGIN
alignment_scores
Quality: 688.00 Length: 124
Ratio: 5.548 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.387
alignment_block:
US-09-327-750D-12 x AA215070
Align seg 1/1 to: AA215070 from: 1 to: 557
1 MetAlaasnValHIsGInGlueAsnGlueMetGlueInProLeuGlnAs 17
166 ATGCCCATGCTCCACCAGGAAACGAGAGATGGAGCAGCCCTCGAGAA 215
17 nelyGlueLuAspArgProValGlyGlyGlyGlyHisGlnProAlaG 34
216 TGGACAGAGACAGCCGCTGGGAGAGATGGAGCCACAGCCTCGTGTG 265
34 lyeAsnAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 315
266 CAAACAG 315
51 GlyGlnAlaArgLeuAlaProAsnPheArgTrrAlaIleProAsnAr 67
316 GGCACAGCTGCCGACCTTCCCTTACTCTCCATGGGCGCCATTCACAC 365
67 glnMetAsnAspGlyLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 84
366 GCAATGAAATGACGGGTGGGTGGAGATGGATGGATGGATGGATGGAT 415
84 etGlueLueMetArgGlyLeuArgGlyLeuArgGlyLeuGlnLeuArg 100
416 TGGAGGACATGAGAGATGCGGAGAAAGCTTATAGGACGTACAGCTGAG 465

51 GlyClnAlaIaArgArgLeuLeuAlaProAsnPheArgTrrAlaIleProAsnAr 67
157 GCGCAGTGGCGACTGGCCCTACCTCCGATGGCGCTTCCACAG 206
67 GCluMetAsnArgLeuGlyGlyAspGlyAspAspMetCiluMetPheM 84
207 GCAGTGAATGACGGGTTCGGTGGAGATGGAGATGATATGGAAATGTCN 256
84 etGluGluMetArgCgluIleArgArgLysLeuMetGluLeuGluLeuArg 100
257 TGGAGGAGATGAGAGAGATCCGGAGAAAGCTTAGGGAGCTACAGCTGAG 306
101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisH1AspHisH1 117
307 AATTGTCTACGATCTCTTATGGGGAGCTGTCTAACCAACACGATCACCA 356
117 aspGluPheCysLeuMetPro 124
357 TCATGAATTCTGCTTATGCGCT 378

117 saspGluPheCysLeuMetPro 124
|||||
357 TGATGAATTCCTGCCTTATGCCT 378

seq_name: qb_est1:AW476468	
seq_documentation_block:	
LOCUS AW476468	628 bp
	mRNA
	EST
	24-FEB-2000

DEFINITION uc76d02.y1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:2937027 5' similar to cb-w3188 OVARIAN GRANULOSA CELL 13.0 KDa PROTEIN HCB7A

(HUMAN);, mRNA sequence.

ACCESSION	AW476468
VERSION	AW476468.1 GI:7046574

RECORDS	SOURCE	ORGANISM
251.	house mouse.	
	Mus musculus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	

REFERENCE
AUTHORS
TITLE
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
1 (bases 1 to 628)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap/>
National Cancer Institute Cancer Genome Anatomy Project (CGAP)

JOURNAL COMMENT
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.

Email: cyapus-remail@nfl.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Cloning distribution: NCI-CCAP clone distribution Information can be
found through the I.M.A.G.E. Consortium website.

www.llnl.gov/bbrp/image/image.htm)

FEATURES
source 1. .628
Location/Qualifiers

```
/organism="Mus musculus"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:2937027"
```

```
/clone_lib="NCI_CGAP_Lu33"  
/tissue_tumor="pooled_lung_tumors"
```

```

/lab_host="DH10B (phage-resistant)"
/note=Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st

```

strand cDNA was prepared from mRNA obtained from pooled lung tumors with a Not I - oligo(dT) primer [5]. TGTACCAATCGAATGGGGAGCGCGCCCTGTTTTTTTTTTTTT 3'). Double-stranded cDNA was ligated to Eco RI adaptors

I and Eco RI sites of the modified p7T3 vector. Library was digested with Not I and cloned into the Not I (Pharmacia), went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo.

101 AaccysleuAargileuMetGlyLeuSerAasnHisAaspHis117
|||||
478 AATGTCTACCATCTTATGGGGAGCTGTCTAACCAACCCAGCATCA 527
117 sAgGluPhcCysLeuMetPro 124
528 TGTGGAATTCCTGCTTATGCTT 549
seq_name: gb_est1:A1152323
seq_documentation_block: 430 bp mRNA EST 30-SEP-1998
LOCUS A1152323
DEFINITION ud79a02.r1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
IMAGE:1477034 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
RD PROTEIN HGR74 (HUMAN); mRNA sequence.
ACCESSION A1152323
KEYWORDS EST:2323.1 GI:368072
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Norris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theisinger, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
The Washington University EST Project
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.lml.gov) for further information.
MGI:94290
Seq primer: -28m3 rev2 ET from Amersham
High quality sequence stop: 419.
FEATURES
source
1..430
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1477034"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="vector: p7730-Pac (Pharmacia) with a modified
polylinker. 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified p773 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 135 a 111 c 113 g 71 t
ORIGIN
alignment_scores:
Quality: 681.00 Length: 124
Ratio: 5.517 Gaps: 0
Percent Similarity: 99.194 Percent Identity: 97.581
alignment_block:
US-09-327-750d-12 x A1152323
Align seg 1/1 to: A1152323 from: 1 to: 430

1 MetAaSnValHisGlnGluAasnGluGluMetGluGlnProLeuGlnAs 17
|||||
54 ATGCCCAATGTCCACCAAGAAAGAGAGATGGAGCAGCCCTGCAGAA 103
17 nGlyGluGluAaspArgProValGlyGlyGlyGlyHisGlnProLeuG 34
|||||
104 TGGACAGGAGACCCCTGTGTGATGGAGTGGAGGCCACCAAGCTGCTG 153
34 LyAasnAasnAasnAasnAasnHisAasnHisAasnHisAasnHisA 50
|||||
154 CAACACACACACACACACACACACACATACCAACACACACACACGA 203
51 GlyGlnAaArgArgLeuAlaProAasnPheArgTTPalaileProAasn 67
|||||
204 GCCCAGGCTCCCGACCTTCCCTTAACCTTCGGATGGGCCATTCCCAAC 253
67 GlnMetAasnAaspGlyLeuGlyGlyAaspGlyAaspAaspMetGlu 84
|||||
254 GCAGATGATGACCGCTTGGGTGGAGATGGAGATGATATGGAATCTCA 303
84 etGluGluMetArgGluLeuArgArgGlyLeuArgGlyLeuGluLeu 100
|||||
304 TGGAGGAGATGAGAGAGATCCGGAGAAAGCTTAGGGAGCTACAGCTGA 353
101 AaccysleuAargileuMetGlyLeuSerAasnHisAaspHis117
|||||
354 AATGTCTACCATCTTATGGGGAGCTGTCTAACCAACCCAGCATCA 403
117 sAgGluPhcCysLeuMetPro 124
404 TGTGGAATTCCTGCTTATGCTT 425
seq_name: gb_est1:AW908751
seq_documentation_block: 503 bp mRNA EST 25-MAY-2000
LOCUS AW908751
DEFINITION uf57a05.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
IMAGE:1515440 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
RD PROTEIN HGR74 (HUMAN); mRNA sequence.
ACCESSION AW908751
KEYWORDS EST:2323.1 GI:8073984
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
NCI-CGAP, Bethesda, Maryland
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.lml.gov) for further information.
MGI:94292
Seq primer: -40Rp from Gibco
High quality sequence stop: 468.
FEATURES
Location/Qualifiers
1..503
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1515440"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/lab_host="mammary gland"
/lab_host="DH10B"
/note="vector: p7730-Pac (Pharmacia) with a modified
polylinker. 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not

```

gb_hvg:LMFCHRH16_01 - 91.00 107.47 1.0e+03 1100000 .! Continuation (2 of 11)
gb-pl:AVTAGUES + 88.50 117.18 299.25 17579 ! L39096 Acetobacter vinel
gb-pl:AVTAGUES + 87.50 127.24 82.39 3000 ! AJ251864 Polypocladium inl
92_256:WMA52491 - 90.12 116.39 256.15 10612 ! AEO04479 Neisseria mening
92_256:WMA52491 - 87.50 94.33 5.0e+03 323861 ! AL162756 Neisseria mening
seq_name: gb_pl:BC003190
seq_documentation_block:
LOCUS BC003190 793 bp mRNA PRI 12-JUL-2001
DEFINITION Homo sapiens, p75NTR-associated cell death executor; ovarian
granulosa cell protein (13KD), clone MGC:802 IMAGE:3357965, mRNA,
complete cds.
ACCESSION BC03190
VERSION BC03190.1
KEYWORDS MGC.
SOURCE GT:131112030
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 793)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (13-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 1A03, Bethesda, MD 20892-2590,
NHL-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cmcgmbs@fghml.nih.gov
REMARK
COMMENT

```

```

Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLML)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
StorSeqSGS.BC.ca
Susanna Chan, Readman Chiu, Ian Boudet, Yaron Butterfield,
Leticia Hsiao, Martin Krzywinski, Reta Kutache, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Lilisa Prabhu, Parvaneh Saeedi, Jacqueline
Schlein, Duane Smialus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Minarada Tsei, Natasha van den Bosch, Jill Vardy,
George Yang, Scott Zyuderduyn, Marco Zatta.

Clone distribution: MGC clone distribution information can be found
through the following URL: http://www.ncbi.nlm.nih.gov/
Series: IPAL place, February 7, 2006
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 7657043
Location/Qualifiers
    1. .793
       /organism="Homo sapiens"
       /db_xref="LocusID:27018"
       /db_xref="taxon:9606"
       /clone_MGC=802 IMAGE:3357965"
       /clone_1lib="NHK-MC-16"
       /lab_host="DH10B-K"
       /note="Vector: pOTB7"
       196. .531
       /product="p75NTR-associated cell death executor: ovarian
granulosa cell protein (l3kb)"
       /protein_id="AA043190.1"
       /translation="MANTHNQEMENPQNGBEDPLGGGSHOPAGNRPCQAERLA
TPNSATIGASVETDPTDGLRPPNRIQNDGGGDDETFMEHRIRKLRELQNCRLANGLSNNHD
PNFNAIPNRIQNDGGGDDETFMEHRIRKLRELQNCRLANGLSNNHD"
       HIDEFCLMP"
225 a 170 c 205 g 193 t

```


MEDLINE 20298829
 REFERENCE 2 (bases 1 to 891)
 AUTHORS Mukai, J., Hachiya, T., Hoshino, S., Kimura, M., Nadano, D., Suvanto, P.,
 Haneoka, T., Li, Y., Irie, S. and Sato, T.
 TITLE Direct Submission
 JOURNAL Submitted (17-SEP-1999) Otolaryngology/Pathology, Columbia
 University, 630 West 168th St., P&S 11-451, New York, NY 10032, USA
 FEATURES
 source /organism="Mus musculus"
 /db_xref="taxon:9606"
 gene 1..891
 CDS 312..647
 /gene="NADE"
 /note="Involved in the common neurotrophin receptor
 p75NTR-mediated signal transduction"
 /codon_start=1
 /product="p75NTR-associated cell death executor"
 /protein_id="AAF75129.1"
 /db_xref="GI:8452894"
 /translation="MAANVHGENEENQPNQEDRPVGGSGHQPAGNRRGQARLLA
 PNFNAIPNRIKNDGGDDDEIFNEIRKLRELQNLCLLWGLSLNHH
 HIDECLM"
 BASE COUNT 251 a 182 c 224 g 234 t
 ORIGIN
 alignment_scores
 Quality: 609.00 Length: 111
 Ratio: 5.486 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block
 US-09-327-750D-13 x AF187064
 Align seg 1/1 to: AF187064 from: 1 to: 891
 1 MetAlaAsnTleHisGlnGluAnGluMetGluGlnProMetGlnAs 17
 312 ATGGCAATATTCACCGAGAAACGAGAGATGGAGCGCTATGCAGAA 361
 17 nglyglulaspargproLeuGlyGlyGlyGlyHisGlnProAlaG 34
 362 TGGAGAGAGAGACCGCTTTGGAGAGAGTGAGGCCACCGCTGCAG 411
 34 lYAsnArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgPhe 50
 412 GAAATCGACGGGACAGCGCTCCGAGCTTCCCTTAATTTCCATGGGCC 461
 51 lIleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspPme 67
 462 ATACCAATAGGCAGATCATATGATGGGATGGTGAGATGGAGATGATAT 511
 67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGlu 84
 512 GGAATATTCAGGAGATGAGAAATCAGAAACCTTAGCGACC 561
 84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGlyLeuSerAsnHis 100
 562 TGCATTTGGAGAAATGTCTGCTATCTTATGGGAGCTCTCTAATCAC 611
 101 HisAspHisAspGluPheCysLeuMetPro 111
 612 CATGACCATCATGATGATTTTCCCTTATGGCT 644
 seq_name: gb_ro:AF187066
 seq_documentation_block:
 LOCUS AF187066 700 bp mRNA 11-JUN-2000
 DEFINITION Mus musculus p75NTR-associated cell death executor (NADE) mRNA,
 complete cds.
 ACCESSION AF187066

AF187066.1 GI:8452897
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 700)
 Mukai, J., Hachiya, T., Shoji-Hoshino, S., Kimura, M. T., Nadano, D.,
 Suvanto, P., Haneoka, T., Li, Y., Irie, S., Greene, L. A. and Sato, T. A.
 NADE, a p75NTR-associated cell death executor, is involved in
 signal transduction mediated by the common neurotrophin receptor
 p75NTR.
 J Biol Chem. 275 (23), 17566-17570 (2000)
 JOURNAL 20298829
 MEDLINE
 REFERENCE 2 (bases 1 to 700)
 AUTHORS Mukai, J., Hachiya, T., Hoshino, S., Kimura, M., Nadano, D., Suvanto, P.,
 Haneoka, T., Li, Y., Irie, S. and Sato, T.
 TITLE Direct Submission
 JOURNAL Submitted (17-SEP-1999) Otolaryngology/Pathology, Columbia
 University, 630 West 168th St., P&S 11-451, New York, NY 10032, USA
 FEATURES
 source /organism="Mus musculus"
 /strain="BALB/c"
 /db_xref="taxon:10090"
 gene 1..700
 CDS 177..551
 /gene="NADE"
 /note="Involved in the common neurotrophin receptor
 p75NTR-mediated signal transduction; NADE"
 /codon_start=1
 /product="p75NTR-associated cell death executor"
 /protein_id="AAF75131.1"
 /db_xref="GI:8452898"
 /translation="MAANVHGENEENQPNQEDRPVGGSGHQPANNNNNHHN
 HIDECLM"
 BASE COUNT 178 a 187 c 203 g 132 t
 ORIGIN
 alignment_scores
 Quality: 567.50 Length: 124
 Ratio: 5.113 Gaps: 1
 Percent Similarity: 89.516 Percent Identity: 83.065
 alignment_block
 US-09-327-750D-13 x AF187066
 Align seg 1/1 to: AF187066 from: 1 to: 700
 1 MetAlaAsnTleHisGlnGluAnGluMetGluGlnProMetGlnAs 17
 177 ATGGCAATATTCACCGAGAAACGAGAGATGGAGCGCTATGCAGAA 226
 17 nglyglulaspargproLeuGlyGlyGlyGlyHisGlnProAlaG 34
 227 TGGAGAGAGAGACCGCTTTGGAGAGAGTGAGGCCACCGCTGCAG 276
 34 lYAsn.....ArgAsp 37
 277 CAAC 326
 38 GlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnAr 54
 327 GCCCAGCTCCGCACTTGGCCCTTAATCTCGATGGCCATTCGCCACAG 376
 54 gGlnIleAsnAspGlyMetGlyGlyAspGlyAspMetGluIlePheM 71
 377 CGCATGATGACGGTTGGGTGGAGATGAGATGATGAAATGTTC 426
 71 etclGluMetArgGluIleArgArgLysLeuArgGluLeuGlnLeuArg 87


```
Ratio: 3.571      Gaps: 4
Percent Similarity: 73.684      Percent Identity: 57.018

alignment_block:
US-09-327-750d-13 x HSV351F8

Align seg 1/1 to: HSV351F8 from: 1 to: 45678

1 MetAlaLeuHleHleGlnGluAnGluMetGluGln...PrometG1 16
|||||
19456 ATGGAATATGTCCTCCCAAGGAAACAAAGTTGTGGAGAGGCCGCCAGTGC 19505
|||||
16 nAnGlyGluGluAspArgProLeuGlyGlyGlyGlyGlyHleGlnProA 33
|||||
19506 GAAT.....GAAGCCCGCTTTAGGAGGTGGTGAATACAGGAGCCG 19549
|||||
33 laGlyAnArgArgGlyGlnAlaArgArgLeuAlaProAnPheArgTTP 49
|||||
19550 GAGGAATATGTAAGGGGTGTGGGCTCCAGTCCCGCGGGTTTGGAGAG 19599
|||||
50 AlAlleArgOnsArgGlnIleAn.....AspGlyMetGlyGlyAspG1 64
|||||
19600 GATGTGCCCATAGCTGTGCGATACCAATGATATAGATAGATGGAGATGG 19649
|||||
64 YAnPAspMetGluIlePheMetGluGluMetArgGluIleArgGlyLys 81
|||||
19650 AGATGATATGGAACGGTTTCATGAGGAGATGAGAGAGCTTAAGGAGGAAA 19699
|||||
81 euArgLeuGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeu 97
|||||
19700 TTAGGAACTTCAGTTGAGGTACAGTCTGCCATCTTTATAGGGGAC... 19746
|||||
98 SerAnHleHisAspHleHisAspGluPheCysLeuMetPro 111
|||||
19747 CCTCTCCACCGATCATCATCATCATCATCATCATCATCATCATCATCAT 19788

seq_name: gb_pr:AK000959

seq_documentation_block:
LOCUS AK000959 1229 bp mRNA PRI 22-FEB-2000
DEFINITION Homo sapiens cDNA FLJ10097 fis, clone HEMBA1002458, weakly similar
to OVIARIAN GRANULOSA CELL 13.0 KD PROTEIN.HGR74.
ACCESSION AK000959.1 GI:7021945
VERSION 1
KEYWORDS oligo capping, f18 (full insert sequence).
SOURCE human
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1229)
AUTHORS Isogai,T., Ota,T., Hayaishi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H.,
Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K.,
Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yanamoto,J.,
Makamatsu,A., Nakamura,T., Negahari,K., Masuhara,I. and Oshima,A.
NEDO human cDNA sequencing project
JOURNAL Nucleic Acids Res. 29(10):1945-1946 (2001)
REFERENCE 2 (bases 1 to 1229)
AUTHORS Isogai,T. and Otsuki,T.
JOURNAL Submitted (16-FEB-2000) to the DDBJ/EMBL/Genbank databases. Takao
Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
Project); Department of Virology, Institute of Medical Science,
University of Tokyo
FEATURES
Location/Qualifiers
```

```
source
1..1229
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1002458"
/dev_stage="embryo, 10 weeks"
/tissue="cloning vector: pME18FLJ"
/notes="unnamed protein product"
191..353
/contig="contig1"
/protid="BA091443.1"
/db_xref="GI:7021945"
/translation="MESKEELANNGENAOQENEGEODAPTONEEZSRHGGGEGG
KFGNIRGRVRLVNPFRWALPNHIEARDDVERFVGOMKIKRKTREOQMRHY
MFQTEPDNHYDFCLIP"
BASE COUNT 351 a 243 c 296 g 339 t
ORIGIN

alignment_scores:
Quality: 277.00 Length: 111
Ratio: 3.298 Gaps: 4
Percent Similarity: 75.676 Percent Identity: 54.955

alignment_block:
US-09-327-750d-13 x AK000959

Align seg 1/1 to: AK000959 from: 1 to: 1229

3 AsnIleHleGlnGluAnGluMetGluGln...PrometGlnAnG1 18
|||||
236 ATGCCCAACAGAAACAGAGCGGAGCGGCCGCCACCGCGATGA 285
|||||
18 YGluGluAspArgProLeuGlyGlyGlyGlyGlyHleGlnProAlaGlyA 35
|||||
286 AGAAGATCCCGCCATTTGGGAGGGGTGAAGCCAGAGCCTGGAGGAA 335
|||||
35 sn...ArgArgGlyGlnAlaArgArgLeuAlaProAnPheArgTTPAla 50
|||||
336 ATATCAGCGGGCGGAGGTAGGCGACTGTCCCTAATTTTCGATGGGCC 385
|||||
51 IleProAnArgGlnIleAnAspGlyMetGlyGlyAspGlyAspAspMc 67
|||||
386 ATACTTATGCGCATATGACACATGAAGCG.....AGAGTGTGTG 429
|||||
67 tGluIlePheMetGluGluMetArgGluIleArgArgGlyLeuArgGluL 84
|||||
430 AGAAGGTTTGTAGGCGCATGTGGAATCAAGAGAAAGACTAGGAGAAC 479
|||||
84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAnHle 100
|||||
480 AGCAGTAGGCGGACTATATGCGGCTTCCAACTCCTGACCTGACACCAT 529
|||||
101 HisAspHleHisAspGluPheCysLeuMetPro 111
|||||
530 TAGGAC.....TTTTCCTCATACCT 550

seq_name: gb_pat:AK100231

seq_documentation_block:
LOCUS AK100231 1364 bp DNA PAT 02-APR-2001
DEFINITION Sequence 32 from Patent WO0119860.
ACCESSION AK100231
VERSION AK100231.1 GI:13539115
KEYWORDS
SOURCE human
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1364)
AUTHORS Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eumetazoa; Echinodermata; Primates; Catarrhini; Hominoidea; Homo.
Tang,Y.T., Hillman,J.L., Yue,H., Reddy,R., Lal,P., Shah,P.,
Azimzai,Z., Baughn,M.R., Lu,D.A., Bandman,O., Shih,L.L. and
```


BASE COUNT 256 a 170 c 221 g 181 t
ORIGIN HSLRAVSTDPFHHDHDFCLMP*

alignment_scores:
Quality: 194.50 Length: 113
Ratio: 2.739 Gaps: 3
Percent Similarity: 62.832 Percent Identity: 43.363

alignment_block:
US-09-327-750D-13 x AF220189 ..

Align seg 1/1 to: AF220189 from: 1 to: 828

```
1 MetAlaAsnIleHisGlnGluAnGluMetGluInProMetGlnAs 17
||| ||| :|||:|||||:||||| :|||: :|||:
206 ATGGAAATGCCAACCAAGAAATGAGAAAGAGCAAGTTCCTAATAA 255
17 nGlyGluGlu...AspArgProLeuGlyGlyGlyGlyHisGlnProA 33
:||||| :||||| :||||| :||||| :||||| :|||
256 AGGGAGGAGCCCTGGCCCTCTTGGATGCTGTGTATCTGCTGCTC 305
33 IscIAspArgArgGly...GlnAlaArgArgLeuAlaProAsnPheArg 48
||||| :||||| :||||| :||||| :||||| :|||
306 GAGGAATCTAGCGGCTTCGCGTTAGCGCCATCCTGCGATATAGA 355
49 TrpAlaIleProAsnArgGlnIleAsnAspGlyMetGlyAspGlyAs 65
||| :||| :||| :||| :||| :|||
356 TGGATATGATGATAGCTTGGAGAACACAGCAAGGATGAGAGAAGA 405
65 pApMetGluIlePheMetGluGluMetArgIuIleArgArgLysLeuA 82
:||||| :|||: :|||: :|||: :|||: :|||:
406 GATATGCGAAGCATGGGAGGAGGTGACACGCTGTGGAAGACCTGA 455
82 TGGIuGluGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSer 98
||||| :||||| :|||: :|||: :|||: :|||
456 GGGAAAGACGTTGAGTCATGCTCGGGCAGTCACACACAG...CCC 502
99 AaHHisAspHisHisAspLupPheCysLeuMetPro 111
||||| :||||| :||||| :||||| :||||| :|||
503 CCTCACCATGACCATCATGATGAGTTTGGCCTTATGCC 541
```

Strd	Orig	zScore	EScore	Len	Documentation
gb_ses2c.N34237	+	609.0	1048.29	3.2e-49	1519
gb_ses2c.B6G3170	+	609.0	1046.94	3.3e-49	335
gb_ses2c.B6G3170	+	609.0	1046.94	3.3e-49	335
gb_ses1c.A1193112	+	609.0	1046.86	3.3e-49	615
gb_ses1c.A1193112	+	609.0	1046.86	3.3e-49	615
gb_ses1c.A576958	+	609.0	1046.79	3.3e-49	620
gb_ses1c.A576958	+	609.0	1046.79	3.3e-49	620
gb_ses1c.A159291106	+	609.0	1046.24	4.1e-49	652
gb_ses1c.A159291106	+	609.0	1046.24	4.1e-49	652
gb_ses1c.B6754809	+	609.0	1046.08	4.2e-49	674
gb_ses1c.B6754809	+	609.0	1046.08	4.2e-49	674
gb_ses1c.AV701506	+	609.0	1045.58	4.5e-49	720
gb_ses1c.B6F098872	+	600.0	1031.77	2.6e-48	586
gb_ses1c.B6F098872	+	600.0	1031.77	2.6e-48	586
gb_ses2c.B6G74807	+	597.0	1032.03	3.4e-48	557
gb_ses2c.B6G74807	+	597.0	1032.03	3.4e-48	557
gb_ses2c.B6B636153	+	597.0	1035.39	6.3e-48	676
gb_ses2c.B6B636153	+	597.0	1035.39	6.3e-48	676
gb_ses2c.B6B6393540	+	596.0	1035.42	5.9e-48	559
gb_ses2c.B6B6393540	+	596.0	1035.42	5.9e-48	559
gb_ses1c.B6252272	+	591.0	1016.67	1.8e-47	656
gb_ses1c.B6252272	+	591.0	1016.67	1.8e-47	656
gb_ses1c.A15929835	+	589.0	1012.27	3.2e-47	636
gb_ses1c.A15929835	+	589.0	1012.27	3.2e-47	636
gb_ses1c.A576541324	+	585.0	1005.92	7.5e-47	603
gb_ses1c.A576541324	+	585.0	1005.92	7.5e-47	603
gb_ses2c.B6G74807	+	584.0	1001.81	1.2e-46	779
gb_ses2c.B6G74807	+	584.0	1001.81	1.2e-46	779
gb_ses2c.W40094	+	583.0	1000.83	1.4e-46	715
gb_ses2c.W40094	+	583.0	1000.83	1.4e-46	715
gb_ses1c.B6709912	+	582.0	1002.41	1.0e-46	498
gb_ses1c.B6709912	+	582.0	1002.41	1.0e-46	498
gb_ses2c.B6718616	+	582.0	1002.15	1.2e-46	438
gb_ses2c.B6718616	+	582.0	1002.15	1.2e-46	438
gb_ses2c.B6W59318	+	582.0	1001.99	1.2e-46	507
gb_ses2c.B6W59318	+	582.0	1001.99	1.2e-46	507
gb_ses2c.B6F98483	+	582.0	1001.68	1.4e-46	536
gb_ses2c.B6F98483	+	582.0	1001.68	1.4e-46	536
gb_ses1c.B6Z58587	+	582.0	1001.43	1.3e-46	542
gb_ses1c.B6Z58587	+	582.0	1001.43	1.3e-46	542
gb_ses1c.AM113584	+	582.0	1001.03	1.3e-46	568
gb_ses1c.AM113584	+	582.0	1001.03	1.3e-46	568
gb_ses1c.AV705889	+	582.0	1000.96	1.4e-46	572
gb_ses1c.AV705889	+	582.0	1000.96	1.4e-46	572
gb_ses2c.B6716261	+	582.0	1000.74	1.4e-46	568
gb_ses2c.B6716261	+	582.0	1000.74	1.4e-46	568
gb_ses2c.B6715325	+	582.0	1000.57	1.4e-46	600
gb_ses2c.B6715325	+	582.0	1000.57	1.4e-46	600
gb_ses1c.B67093128	+	582.0	1000.43	1.5e-46	610
gb_ses1c.B67093128	+	582.0	1000.43	1.5e-46	610
gb_ses2c.B67053415	+	582.0	1000.36	1.5e-46	615
gb_ses2c.B67053415	+	582.0	1000.36	1.5e-46	615
gb_ses1c.AV732694	+	582.0	999.81	1.6e-46	656
gb_ses1c.AV732694	+	582.0	999.81	1.6e-46	656
gb_ses2c.B6F731092	+	582.0	999.65	1.6e-46	669
gb_ses2c.B6F731092	+	582.0	999.65	1.6e-46	669
gb_ses1c.B6773740	+	582.0	998.29	1.9e-46	692
gb_ses1c.B6773740	+	582.0			

Allan seq 1/1 to: N34237 from

Align seg 1/1 to: N34237 from: 1 to: 519

human

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml
Seq primer: -400P from Glibco
High quality sequence stop: 422.

FEATURES

Location/Qualifiers
1..612
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CGAP,lym12"
/tissue_type="lymphoma, follicular mixed small and large
cell"
/lab_host="DH108"
/note="Organ: lymph node; Vector: pCMV-SPORT6; Site:1:
Salt; Site:2: NotI; Cloned unidirectionally. Primer:
oligo dh. Average insert size 1.25 kb. Life Technologies
catalog #: 11547-015"
BASE COUNT 169 a 145 c 120 g 178 t
ORIGIN

alignment_scores:
Quality: 609.00 Length: 111
Ratio: 5.486 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-327-750D-13 x AW512400/rev ..

Align seg 1/1 to reverse of: AW512400 from: 1 to: 612

1 MetAlaAGNleHieHicIngluAnsgluMetGluInProMetGlnAs 17
581 ATGCCAAATATTCACGAGAAACGAGAGATGGAGCCATATGCAGAA 532
17 nGlyGluGluAspArgProLeuGlyGlyGlyGlyHisGlnProHla 34
531 TGGAGAGGAGACCCCTTTGGGAGGAGGTGAAGCCACCGAGCTGAG 482
34 lYAsnArgArgGlyGlnAlaArgArgLeuAlaProAsnPhaArgTrpAla 50
481 GAAATCGAGGGGACAGCCCTCGCCGACTTGCCTTAATTTTCGATGGCC 432
51 lIleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspPme 67
431 ATACCCATAGCCAGATCAATATGGATGGGTGGGATGGAGATGATAT 382
67 tGluIlePheMetGluGluMetArgGluIleArgArgLeuArgGlu 84
381 GCAATATTCATGAGGAGATGAGAGAAATCAGAGAAACTTAGGGAGC 332
84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHis 100
331 TGCAGTTGAGGAATTTGTCGCTATCTTATGGGGAGCTCTCTAATCAC 282
101 HisAspHisHisAspGluPheCysLeuMetPro 111
281 CATGACCATCATGATGAATTTGGCTTATGCTT 249

seq_name: gb_est1.A1193112

seq_documentation_block:
LOCUS A1193112 615 bp mRNA EST
DEFINITION Q65909.x1 Soares_fetal_lung_NDHL19W Homo sapiens
IMAGE:1744288.3' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
KD PROTEIN HCR74 (HUMAN); mRNA sequence.

ACCESSION A1193112
VERSION A1193112.1 GI:3744321
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 615)
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Primates; Catarrhini; Homnidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cygaps@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (image.llnl.gov) for further information.
Insert length: 532 bp
Seq primer: -400P from Glibco
High quality sequence stop: 445.

FEATURES

Location/Qualifiers
1..615

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Soares_fetal_lung_NDHL19W"
/dev_stage="19 weeks"
/lab_host="DH108 (ampicillin resistant)"
/note="Organ: lung; Vector: p773D (Pharmacia) with a
modified EcoRI linker; Site:1: NotI; Site:2: EcoRI
linker; Cloned unidirectionally. Primer:
15'-TGATCAACACGAGTGGAGCGCCGCAATTTTATTTT-3';
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p773 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Ronaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NDHL19W."

BASE COUNT 169 a 145 c 121 g 180 t
ORIGIN

alignment_scores:

Quality: 609.00 Length: 111
Ratio: 5.486 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-327-750D-13 x A1193112/rev ..

Align seg 1/1 to reverse of: A1193112 from: 1 to: 615

1 MetAlaAGNleHieHicIngluAnsgluMetGluInProMetGlnAs 17
584 ATGCCAAATATTCACGAGAAACGAGAGATGGAGCCATATGCAGAA 535
17 nGlyGluGluAspArgProLeuGlyGlyGlyGlyHisGlnProHla 34
534 TGGAGAGGAGACCCCTTTGGGAGGAGGTGAAGCCACCGAGCTGAG 485
34 lYAsnArgArgGlyGlnAlaArgArgLeuAlaProAsnPhaArgTrpAla 50
484 GAAATCGAGGGGACAGCCCTCGCCGACTTGCCTTAATTTTCGATGGCC 435
51 lIleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspPme 67
434 ATACCCATAGCCAGATCAATATGGATGGGTGGGATGGAGATGATAT 385
67 tGluIlePheMetGluGluMetArgGluIleArgArgLeuArgGlu 84
384 GCAATATTCATGAGGAGATGAGAGAAATCAGAGAAACTTAGGGAGC 335
84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHis 100
334 TGCAGTTGAGGAATTTGTCGCTATCTTATGGGGAGCTCTCTAATCAC 285
101 HisAspHisHisAspGluPheCysLeuMetPro 111

DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/IMG at:
 image.lnl.gov/image/hcm1/resources.shtml
 seq primer: 40UP from Gibco
 High quality sequence stop: 422.

FEATURES

Location/Qualifiers
 1..612
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2849369"
 /clone_lib="NCI-CGAP-Lym12"
 /tissue_type="lymphoma, follicular mixed small and large
 cell"
 /lab_host="DH10B"
 /note="Organ: lymph node; Vector: pCMV-SPORT6; Site: 1:
 Note: Site 2: Not; Cloned unidirectionally. Primer:
 Oligo an: 11547-015"
 BASE COUNT 169 a 145 c 120 g 178 t
 ORIGIN

alignment_scores
 Quality: 609.00 Length: 111
 Ratio: 5.486 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-327-750D-13 x AN512400/rev ..

Align seg 1/1 to reverse of: AN512400 from: 1 to: 612

1 MetAlaAsnIleHisGlnGluAnGluMetGluGlnProMetGlnAs 17
 581 ATGCCAAATATTCACCAAGAAACAGAGATGGAGCGCTATGCAGAA 532
 17 nGlyGluGluAspArgProLeuGlyGlyGlyHisGlnProAlaG 34
 531 TGGAGAGAGACCGCCCTTTGGGAGAGGTGAAGGCCACCGCTGCAG 482
 34 lYAspArgGlyGlnIleAlaArgGlyLeuAlaProAsnPheArgTrpAla 50
 481 GAAATCGACGGGACAGCGCTGCCCTTATTTTCGATGGGCC 432
 51 lIleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspAspMe 67
 431 ATACCAATAGGCAGATCAATGATGGATGGGTGGGATGGAGATGATAT 382
 67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84
 381 GGAATATTCGAGGAGATGAGAGAAATCAGAGAAACCTTAGGAGC 332
 84 euGluLeuArgAsnCysLeuArgGlyLeuMetGlyGlyLeuSerAsnHis 100
 331 TGCAGTTCAGGAATTCCTGCGTATCTTATGGGGAGCTCTCTAATCAC 282
 101 HisAspHisHisAspGluPheCysLeuMetPro 111
 281 CATGACCATCATGATCAATTTTGCCTTATGCCT 249

seq_name: gb_est1: A1193112

seq documentation block:
 LOCUS A1193112 615 bp mRNA EST 29-OCT-1998
 DEFINITION Q66309.1 Soares fetal lung NBHL19W Homo sapiens CD84 clone
 IMAGE:1744288 3' similar to GB-M38188 OVARIAN GRANULOSA CELL 13.0
 KD PROTEIN HGR74 (HUMAN); mRNA sequence.

ACCESSION A1193112
 VERSION A1193112.1 GI:374321
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 615)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

REFERENCE

AUTHORS
 TITLE

JOURNAL
 COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: rstraus@fmail.nih.gov

This is a copyright-free through LNL; contact the
 IMAGE Consortium (info@img.llnl.gov) for further information.
 Insert Length: 845 Std Error: 0.00

seq primer: 40UP from Gibco

High quality sequence stop: 445.

FEATURES

Location/Qualifiers
 1..615
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1744288"
 /clone_lib="Sources_fetal_lung_NBHL19W"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: p773D (Pharmacia) with a
 modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer
 [5'-TGTTACCAATCTGAAGTGGAGCGCCCAATTTTTTTTTTT-3']
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified p773 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M. Patricia Bonaldo. This library was constructed
 from the same fetus as the fetal heart library, Soares
 fetal heart NBHL19W.

BASE COUNT 169 a 145 c 121 g 180 t

alignment_scores:

Quality: 609.00 Length: 111

Ratio: 5.486 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-327-750D-13 x A1193112/rev ..

Align seg 1/1 to reverse of: A1193112 from: 1 to: 615

1 MetAlaAsnIleHisGlnGluAnGluMetGluGlnProMetGlnAs 17
 584 ATGCCAAATATTCACCAAGAAACAGAGATGGAGCGCTATGCAGAA 535
 17 nGlyGluGluAspArgProLeuGlyGlyGlyHisGlnProAlaG 34
 534 TGGAGAGAGACCGCCCTTTGGGAGAGGTGAAGGCCACCGCTGCAG 485
 34 lYAspArgGlyGlnIleAlaArgGlyLeuAlaProAsnPheArgTrpAla 50
 484 GAAATCGACGGGACAGCGCTGCCCTTATTTTCGATGGGCC 435
 51 lIleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspAspMe 67
 434 ATACCAATAGGCAGATCAATGATGGATGGGTGGGATGGAGATGATAT 385
 67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84
 384 GGAATATTCGAGGAGATGAGAGAAATCAGAGAAACCTTAGGAGC 335
 84 euGluLeuArgAsnCysLeuArgGlyLeuMetGlyGlyLeuSerAsnHis 100
 334 TGCAGTTCAGGAATTCCTGCGTATCTTATGGGGAGCTCTCTAATCAC 285
 101 HisAspHisHisAspGluPheCysLeuMetPro 111

```

34  lyAsnAaRGgLyGlnlaAaRGaLeuAaProAsnPhaATrpAla 50
502  GAAATCGACGGGAGAGCTGCGGACTTGCCTTAATTTCGATGGGGCC 453
51  lleProAsnAaRGlnleAsnApgLyMeGlyGlyAspGlyAspAspMe 67
452  ATACCCATATGACAGATCATGATGGATGGGTGGGTGGAGATGATAT 403
67  tGluilePheMetGlUluMeAaRGlnleAaRGaLeuAaRGgLUl 84
402  GGAATATTATTCATGAGGAGATGAGAGAAATCAGAAGAAACATTAGGAGC 353
84  euGlnleAaRGaancCysLeuAaRGilleLeuMetOlyGluLeuSerAanHls 100
352  TCGATATTCAGAGAAATATGTCGTATATCTTATGAGGGAGAGCTCTTAATAC 303
101  HAspAaPheHleAspGluPheCysLeuMetpro 111
302  CATGACCATCATGATGAATTTTGCCTTATGCTT 270

seq_name: gb_est1:A1929106

seq_documentation_block: 652 bp mRNA EST 23-AUG-1999
A1929106
aues010.yl schneider fetal brain 00004 Homo sapiens cDNA clone
IMAGE:25196115, similar to gb:438188 OVARIAN GRANULOSA CELL 13.0
A1929106.1 GR74 (HUMAN), mRNA sequence.
A1929106.1 GI:5665070
EST.
human.
HOMO SAPIENS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 662)
Miller, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S.,
Krizman, D., Kucaba, J., Lacey, K., Se, N., Lennon, G., Marra, M., Martin,
White, Y., Wylie, T., Waterston, R., and Wilson, R.
WashU-NCI human EST project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LML; contact the
LML (lml@wustl.wustl.edu) for more information.
Seq primer: -40BP from Gibco
High quality sequence stop: 437.
Location/Qualifiers
1. .662
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2519611"
/clone_lib="Schneider fetal brain 00004"
/tissue="brain"
/tissue_type="frontal lobe"
/dev_stage="5 months post-conception"
/lab_host="DH108"
/note="Organ: brain; Vector: pBluescript SK (Stratagene);
Site:1: Sct1; Site:2: Xho1; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
5'-GAGCAGAGAGAGCTCAGAGATCTTATTAATTAATTCCTCCCTCCCTCC-3'
3' adaptor:
5'-GAGCAGAGAGAGCTCAGAGATCTTATTAATTAATTCCTCCCTCCCTCC-3'.
The library was
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
cDNA."
FEATURES
source

```

enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LNCIB-Area
Science Park, Trieste, Italy). 2 others

BASE COUNT 178 a 143 c 169 g 170 t
ORIGIN

alignment_scores: Quality: 609.00 Length: 111
Mismatch: 586 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-327-750D-13 x A1929106

Align seg 1/1 to: A1929106 from: 1 to: 662

```

1 MetaZasnlleHISglngluAnsgluMetGlulnPrometGlnAs 17
|||||
96 ATGCCAAATATTCACCAAGAAACCGAGAGATGCGAGCCATGCGAGA 145
|||||
17 nglYglulueAspArgProLeuGlyGlyGlyHisGlnProAlag 34
|||||
146 TGGAGAGAGACCCGCTTTGGAGGAGGTGAAGCCACCCGCTGCGAG 195
|||||
34 lyAsnAArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTIPala 50
|||||
196 GAATTCAGCGGGACAGCGCTGCGGACTTGCCTTAATTTTCGATGGGCC 245
|||||
51 IleProAsnAArgGlnleAsnAspGlyMetGlyGlyAspGlyAspPhe 67
|||||
246 ATCCCAATAGCGAGATCATGATGGATGGGTGGAGATGAGATGATAT 295
|||||
67 tglulilePheMetGlulMetArgGlulileArgGlyLeuAsglul 84
|||||
296 GGAATATTCATGGAGAGATGAGAGAAATCAGAGAAACCTTAGGGAGC 345
|||||
84 euGlnleuArgAsnCysLeuArgIleLeuMetGlyGlyLeuSerAsnHis 100
|||||
346 TGCAGTTGAGGAATTCCTGCGTATCTTATGGGGAGCTCTCTAATCAC 395
|||||
101 HisAspHisAspGluPheCysLeuMetPro 111
|||||
396 CATGACCATCATGATGATATTTTCCTTATGCTT 428

```

seq_name: gb_est2:BG745809

seq_documentation_block:
LOCUS BG745809 674 bp mRNA EST 15-MAY-2001
DEFINITION 60272395171 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:4850293 3',
mRNA sequence.

ACCESSION BG745809
VERSION BG745809.1 GI:14056462

KEYWORDS
EST.
Homo sapiens
ORGANISM

REFERENCE
1 (bases 1 to 674)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS
NIR-MGC http://mgc.nci.nih.gov/
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. Mark Watson
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
Source: Genes, Inc.
Cloned and distributed by: Genes, Inc.
Cloned through the I.M.A.G.E. Consortium (LNL) at:
http://image.lnl.gov
Plate: LCM1691 row: e column: 14
High quality sequence stop: 674.

FEATURES
source

Location/Qualifiers
1..674
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4850293"
/clone_lib="NIH_MGC_113"
/lab_host="DH10B (phage-resistant)"
/note="Organ: spleen; Vector: pOT79; Site:1: XhoI; Site:2:
KcoRI; cDNA made by oligo-dT priming. Directionally cloned
into pGEM-3(+). All sites including the 5' and 3' UTRs
GGCAGAG(G) library constructed by Ling Hong at the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

BASE COUNT 167 a 186 c 142 g 179 t
ORIGIN

alignment_scores: Quality: 609.00 Length: 111
Mismatch: 586 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-327-750D-13 x BG745809/rev

Align seg 1/1 to reverse of: BG745809 from: 1 to: 674

```

1 MetaZasnlleHISglngluAnsgluMetGlulnPrometGlnAs 17
|||||
529 ATGCCAAATATTCACCAAGAAACCGAGAGATGCGAGCCATGCGAGA 480
|||||
17 nglYglulueAspArgProLeuGlyGlyGlyHisGlnProAlag 34
|||||
479 TGGAGAGAGACCCGCTTTGGAGGAGGTGAAGCCACCCGCTGCGAG 430
|||||
34 lyAsnAArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTIPala 50
|||||
429 GAATTCAGCGGGACAGCGCTGCGGACTTGCCTTAATTTTCGATGGGCC 380
|||||
51 IleProAsnAArgGlnleAsnAspGlyMetGlyGlyAspGlyAspPhe 67
|||||
379 ATCCCAATAGCGAGATCATGATGGATGGGTGGAGATGAGATGATAT 330
|||||
67 tglulilePheMetGlulMetArgGlulileArgGlyLeuAsglul 84
|||||
329 GGAATATTCATGGAGAGATGAGAGAAATCAGAGAAACCTTAGGGAGC 280
|||||
84 euGlnleuArgAsnCysLeuArgIleLeuMetGlyGlyLeuSerAsnHis 100
|||||
279 TGCAGTTGAGGAATTCCTGCGTATCTTATGGGGAGCTCTCTAATCAC 230
|||||
101 HisAspHisAspGluPheCysLeuMetPro 111
|||||
229 CATGACCATCATGATGATATTTTCCTTATGCTT 197

```

seq_name: gb_est1:AV701506

seq_documentation_block:
LOCUS AV701506 720 bp mRNA EST 08-OCT-2000
DEFINITION AV701506 ADB Homo sapiens cDNA clone ADBBBH12 5', mRNA sequence.
ACCESSION AV701506
VERSION AV701506.1 GI:10717836

KEYWORDS
EST.
Homo sapiens
ORGANISM

REFERENCE
1 (bases 1 to 720)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS
Peng, Y., Song, H., Huang, Q., Huang, C., Gu, Y., Yang, Y., Gao, G., Xiao
H., Xu, X., Li, N., Qian, B., Liu, F., Ou, J., Gao, X., Cheng, Z., Xu, Z.,
Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu

```

84 euGlNLeuAArgAsnCysLeuAArgileLeuMetGlyGluLeuSerAanHis 100
|||||
336 TCAGATTCAGGAAATTCCTGCGGTATCTCTATGGGGAGCTCTCTAATCAC 287
|||||
101 H1A9PH1SH1A9SPGluPheCysLeuMetPro 111
|||||
286 CATGACCATCATGAGGAAATTTGGCTTATGCTT 254
seq_name: gb_est2:BG482054

seq_documentation_block:
LOCUS BG482054 557 bp mRNA EST 21-MAR-2001
DEFINITION 602527934F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:465153 5',
mRNA sequence.
ACCESSION BG482054
VERSION BG482054.1 GI:13414333
KEYWORDS EST.
SOURCE Human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 557)
AUTHORS NIH-MGC http://mhc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing: InCyte Genomics, Inc.
Cloning Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1434 row: d column: 10
High quality sequence stop: 556.
FEATURES
source
1. 557
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_21"
/clone_image="465153"
/lab_host="DH10B"
/notes="Organ: placenta; vector: pTZ19; Site: 1: XhoI;
Site 2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT 157 a 126 c 166 g 108 t
ORIGIN

alignment_scores:
Quality: 597.00 Length: 111
Ratio: 5.427 Gaps: 0
Percent Similarity: 99.099 Percent Identity: 99.099

alignment_block:
US-09-327-750D-13 x BG482054
Align seg 1/1 to: BG482054 from: 1 to: 557
```

```

1 MetAlaAsnIleHisGlnGluAnGluMetGluGlnProMetGlnAs 17
|||||
185 ATGGCAATTCACGAGAAACGAGAGATGGACGCTATGCGAA 234
|||||
17 nGlyGluGluAsPArgProLeuGlyGlyGlyHisGlnProAlaG 34
|||||
235 TGGAGAGAGACGCCCTTGGGAGAGGTGAAGCCACCGCTCGAG 284
|||||
```

```

34 LYAsnATGATGGLGlnAlaArgATGLeuAlaProAsnPheATGTTpAla 50
|||||
285 GAAATCGACCGGGGACAGCTCGCGACTTCGCCCTTAATTTCCATGGGCC 334
|||||
51 fileProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspPme 67
|||||
335 ATACCCATAGGACAGATCAATCATGGATGGGTGGAGATGGAGATGATAT 384
|||||
67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84
|||||
385 GGAATATTCATGGAGAGATGAGAGAAATCAGAGAAACTTAGGGAGC 434
|||||
84 euGlNLeuAArgAsnCysLeuAArgileLeuMetGlyGluLeuSerAanHis 100
|||||
435 TGCAGTTGAGGAATTCCTGCGGTATCTCTATGGGGAGCTCTCTAATCAC 484
|||||
101 H1A9PH1SH1A9SPGluPheCysLeuMetPro 111
|||||
485 CATGACCATCATGAGGAAATTTGGCTTATGCTT 517
seq_name: gb_est2:BG705685

seq_documentation_block:
LOCUS BG705685 567 bp mRNA EST 07-MAY-2001
DEFINITION 602668910F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4791506 5',
mRNA sequence.
ACCESSION BG705685
VERSION BG705685.1 GI:13980274
KEYWORDS EST.
SOURCE Human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 567)
AUTHORS NIH-MGC http://mhc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshlyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: InCyte Genomics, Inc.
Cloning Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM10658 row: d column: 03
High quality sequence stop: 567.
FEATURES
Location/Qualifiers
source
1. 567
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4791506"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pBlueScriptR (modified
pBlueScriptR); Site: 1: BamHI; Site 2: XbaI;
1: Oligo-dT primed using primer 5'-TTTATTTATTTATTTN-3',
size-selected for average insert size 2.3 kb and
normalized to 80% 5'. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
```

```

BASE COUNT 160 a 130 c 168 g 109 t
ORIGIN

alignment_scores:
Quality: 597.00 Length: 111
```

Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/getchm12.pl?tl=PM04t2-PM0-HT0911-080201-012-h10t3-2001-02-08&t=1)
Seq primer: puc 18 forward
High quality sequence coverage: 550.

FEATURES

1..550
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0911"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORS2ES PCR (U.S. Patents Patent application
No.96/16 to Ludwig Institute for Cancer Research)
No.96/16 to Ludwig Institute for Cancer Research
Tissue mRNA and cDNA amplification were performed under
low stringency conditions."
134 a 113 c 137 g 146 t

BASE COUNT

ORIGIN

alignment_scores:
Quality: 596.00 Length: 112
Ratio: 5.369 Gaps: 1
Percent Similarity: 99.107 Percent identity: 99.107

alignment_block:

US-09-327-750D-13 x BG9993540

Align seg 1/1 to: BG9993540 from: 1 to: 550

1 MetAlaAsnIleHisGlnGluAsnGlnGluMetGlnGlnProMetGlnAs 17
|||||
48 ATGGCAATATTATCCACGAGAAACGAGAGATGGAGAGCTATGCAGAA 97
17 nGlyGluGluAspArgProLeuGlyGlyGlyGlyHisGlnProAlag 34
|||||
98 TGGAG 147
34 IyAsnArgArgGlyGlnAlaArgArgLeuAlaProAsnPhaArgTpaLa 50
|||||
148 GAATCGACGGGACAGGCTCGCCACTTGCCTTAATTTTCGATGGCC 197
50 AtleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspM 67
|||||
198 CATACCAATAGCAGATCAATGATGGATGGGTGGAGATGGAGATGATA 247
67 etGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGlu 83
|||||
248 TGGAAATATTATCGAGAGAGATGAGAGAAATCAGAGAAACTTAGGGAG 297
84 LeuGlnLeuArgAsnCysLeuArgGlyLeuMetGlyGlyLeuSerAsnH 100
|||||
298 CTGCAGTTGAGGAATTTCTGCGTATCTTATGGGGAGCTCTCTAATCA 347
100 sHAspHisHisAspGluPheCysLeuMetPro 111
|||||
348 CCATGACCATCATGATGAATTTTCCTTATGCTTATGCTT 381

seq_name: 9b_gst1:BE252272

seq_documentation_block:

LOCUS BE252272 559 bp mRNA EST 13-JUL-2000
DEFINITION 80114115F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:3354819 5',
ORIGIN BE252272
ACCESSION BE252272
VERSION BE252272.1 GI:9122409
KEYWORDS EST.

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 559)

NIH-MGC http://mhc.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Contact: Robert Strausberg, Ph.D.

Email: csa@bbs.fda.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: L2M157 row: n column: 04

High quality sequence coverage: 559.

FEATURES

source

1..559

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="NIH_MGC_16"

/tissue_type="retinoblastoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:

ScorI; cDNA made by oligo-dt priming. Directionally

cloned into EcoRI/XhoI sites using the following

adaptor: EcoRI/NotI. Library constructed by Ling Hong

at the Laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH-MGC Library."

157 a 126 c 166 g 110 t

BASE COUNT

ORIGIN

alignment_scores:

Quality: 591.00 Length: 111

Ratio: 5.422 Gaps: 0

Percent Similarity: 98.198 Percent identity: 98.198

alignment_block:

US-09-327-750D-13 x BE252272

Align seg 1/1 to: BE252272 from: 1 to: 559

1 MetAlaAsnIleHisGlnGluAsnGlnGluMetGlnGlnProMetGlnAs 17
|||||
177 ATGGCAATATTATCCACGAGAAACGAGAGATGGAGAGCTATGCAGAA 226
17 nGlyGluGluAspArgProLeuGlyGlyGlyGlyHisGlnProAlag 34
|||||
227 TGGAG 276
34 IyAsnArgArgGlyGlnAlaArgArgLeuAlaProAsnPhaArgTpaLa 50
|||||
277 GAATCGACGGGACAGGCTCGCCACTTGCCTTAATTTTCGATGGCC 326
51 IleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspM 67
|||||
327 ATACCAATAGCAGATCAATGATGGATGGGTGGAGATGGAGATGATAT 376
67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84
|||||
377 GGAAATATTATCGAGAGAGATGAGAGAAATCAGAGAAACTTAGGGAGC 426
84 eGlnGlnLeuArgAsnCysLeuArgGlyLeuMetGlyGlyLeuSerAsnH 100
|||||
427 TGCAGTTGAGGAATTTCTGCGTATCTTATGGGGAGCTCTCTAATCAC 476
101 HisAspHisHisAspGluPheCysLeuMetPro 111

ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: alternative 3' end
INDIVIDUAL ISOLATE: starting at nt. 3012 of SEQ ID NO:9
FEATURE:
NAME/KEY: CDS
LOCATION: 2..1021
PCT-US96-02331-12

alignment_scores:
Quality: 82.00 Length: 125
Ratio: 1.302 Gaps: 4
Percent Similarity: 50.400 Percent Identity: 20.800

alignment_block:
US-09-327-750D-13 x PCT-US96-02331-12

Align seg 1/1 to: PCT-US96-02331-12 from: 1 to: 1244

5 HisGlnGluAsnGluGluMetGluGlnProMetGlnAsnGlyGluGluAs 21
|||||
26 CAGGACCGTCTCCGATCGGAACCTGCGAATGCAATCATCGGAGCAGCA 75
|||||
21 p.....ArgProL 24
|||
76 CGACGATCCAGCGCTTATCGAGGAGCTGTTGCGATCACGTTCTGTGAGA 125
|||||
24 euGlyGlyGluGluGluHisGlnProAlaGlyAsnArgArgGlyGlnAla 40
|||||
126 TCGAGGCGGGGAGTACGACGATCCGAGGAGGATGAAGAGGAGCAGCTAC 175
|||||
41 ArgArgLeuAlaProAsnPheArgTrpAlaIle..... 51
|||||
176 CATGCCACACCGCCCAAGTACAGACGGCTGTGTTTATGCTCTCCGCA 225
|||||
52 .ProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspMetG 68
|||||
226 TCGGATGAAGAGCGGCTCCGATCGGATCGGAT..... 262
|||||
68 LuIlePheMetGluGluMetArgGluIleArgGlyLeuArgGlyLeu 84
|||||
263 ..ATCTATGTTGATGCGCGCTACAAATTGCGAGTACAAAGTACAAAGAGCTC 310
|||||
85 GlnLeuArgAsnCysLeuArgIle.....LeuMetGlyGly 96
|||||
311 AACATGCGCGCAACATACGATCGATCGAGTCCGACGACGACATGATGCCA 360
|||||
96 uLeuSerAsnHisHisAspHisHis 104
|||||
361 CTATTCCCGCATCATCGGACCAT 385
|||||

seq_name: /usr2_6/ptodate/2/ins/PCTUS_COMB.seq:PCT-US96-02331-14
seq_documentation_block:
SEQUENCE INFORMATION:
GENERAL INFORMATION:
APPLICANT: The Board of Trustees of the Leland Stanford Junior
APPLICANT: University
TITLE OF INVENTION: Board of Regents, The University of Texas System
TITLE OF INVENTION: Methods and Compositions for Altering
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/02331

FILING DATE: 09-FEB-1996

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/386,495

FILING DATE: 10-FEB-1995

ATTORNEY/AGENT INFORMATION:

NAME: Sholtz, Charles K.

REGISTRATION NUMBER: 38,615

TELEPHONE: (415) 324-0880

TELEFAX: (415) 324-0950

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 4255 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: cDNA to mRNA

HYPERMETHYL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: fruitless transcript in Fig. 7E

FEATURE:

NAME/KEY: CDS

LOCATION: 1507..4032

PCT-US96-02331-14

alignment_scores:

Quality: 82.00 Length: 125

Ratio: 1.302 Gaps: 4

Percent Similarity: 50.400 Percent Identity: 20.800

alignment_block:

US-09-327-750D-13 x PCT-US96-02331-14

Align seg 1/1 to: PCT-US96-02331-14 from: 1 to: 4255

5 HisGlnGluAsnGluGluMetGluGlnProMetGlnAsnGlyGluGluAs 21

|||||

3037 CAGGACCGTCTCCGATCGGAACCTGCGAATGCAATCATCGGAGCAGCA 3086

|||||

21 p.....ArgProL 24

|||

3087 CGACGATCCAGCGCTTATCGAGGAGCTGTTGCGATCACGTTCTGTGAGA 3136

|||||

24 euGlyGlyGluGluGluHisGlnProAlaGlyAsnArgArgGlyGlnAla 40

|||||

3137 TCGAAGCGGGGAGTACGACGATCCGAGGAGGATGAAGAGGAGCAGCTAC 3186

|||||

41 ArgArgLeuAlaProAsnPheArgTrpAlaIle..... 51

|||||

3187 CATGCCACACCGCCCAAGTACAGACGGCTGTGTTTATGCTCTCCGCA 3236

|||||

52 .ProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspMetG 68

|||||

3237 TCGGATGAAGAGCGGCTCCGATCGGATCGGAT..... 3273

|||||

68 LuIlePheMetGluGluMetArgGluIleArgGlyLeuArgGlyLeu 84

|||||

3274 ..ATCTATGTTGATGCGGCTACAAATTGCGAGTACAAAGTCAAGGAGCTC 3321

|||||

85 GlnLeuArgAsnCysLeuArgIle.....LeuMetGlyGly 96

|||||

3322 AACATGCGCGCAACATACGATCGAGTCCGACGACGACATGATGCCA 3371

|||||

96 uLeuSerAsnHisHisAspHisHis 104

|||||

3372 CTATTCCCGCATCATCGGACCAT 3396

|||||

```

978 AAAACGAGGCTCCCAAGGCTACAACTTCAGGTTTGCACAGTACTCAA 1027
994 MetGlyGluLeuSer..... 98
1028 ATGCGAATATGCGAGTACAGTACCGACCCCTCTCAGAGCGTTTGGCATCC 1077
99 AsnHisHisAspH 103
1078 GCTTCGACGCTGCTGATACGGGATCTGGCAGGTTCACATCATCCCG 1127
103 IsrHisAspGluPheCys 108
1128 ACATCATCAGCTCTCT 1144
seq_name: /cgn2_6/ptodate/2/lna/backfiles1.seq:5215881-2
seq_documentation_block:
; Patent: 5215881
; APPLICANT: CHEUNG, ANDREW K.
; TITLE OF INVENTION: PSEUDORABIES DIAGNOSIS PROBES
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/537,855
; FILING DATE: 13-JUN-1990
; SEQ ID NO:2
; LENGTH: 2721
5215881-2

alignment_scores:
Quality: 77.50 Length: 74
Ratio: 1.685 Gaps: 5
Percent Similarity: 62.162 Percent Identity: 35.135
alignment_block:
US-09-327-750D-13 x 5215881-2
Align seg 1/1 to: 5215881-2 from: 1 to: 2721
5 HisGlnGluAnGluMetGluGlnProMet...GlnAsnGlyGluG 20
626 CACTCTGAATAAATCTCTTTCTCACCCTGGAGAGGAGGAGAGG 675
20 uAspArgProLeuGlyGlyGlyHis.....GlnProA 33
676 GGACCGGGGACCGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 725
33 laglyAsn ArgArgGlyGlnAlaArgGlnLeuAlaProAsnPheArgTr 49
726 CGAGCCCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC.....CACACCCCGG 769
49 palAlaProAsnArgGlnIleAsn.....ArgGlyMetGlyG 62
770 CCACCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 819
62 lyAspGlyAspAspMetGlu 68
820 GAGAGGAGGAGGAGGAGGAG 839
seq_name: /cgn2_6/ptodate/2/lna/5A_COMB.seq:US-07-945-283-1
seq_documentation_block:
; Sequence: 1, 5215881
; Patent: 5215881
; APPLICATION US/07945283
; GENERAL INFORMATION:
; APPLICANT: Cheung, Andrew K.
; APPLICANT: Wesley, Ronald D.
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
; TITLE OF INVENTION: Involving the EP0 and LLT Genes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:

```

```

ADDRESSES: Curtis P. Ribando
STREET: 1815 NO. 5352596th University Street
CITY: Peoria
STATE: IL
COUNTRY: USA
ZIP: 61604
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
SOFTWARE: PC-DOS/MS-DOS
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,283
FILING DATE: 19920911
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Ribando, Curtis P
REGISTRATION NUMBER: 27976
TELEPHONE: 309-685-1401 ext.513
TELEFAX: 309-685-1428
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8438 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudorabies virus
FEATURES:
NAME/KEY: CDS
LOCATION: 622...6495
FEATURE:
NAME/KEY: variation
LOCATION: replace(1099, "g")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1267, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1381, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1566, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(7010, "g")
US-07-945-283-1

alignment_scores:
Quality: 77.50 Length: 74
Ratio: 1.685 Gaps: 5
Percent Similarity: 62.162 Percent Identity: 35.135
alignment_block:
US-09-327-750D-13 x US-07-945-283-1
Align seg 1/1 to: US-07-945-283-1 from: 1 to: 8438
5 HisGlnGluAnGluMetGluGlnProMet...GlnAsnGlyGluG 20
1669 CACTCTGAATAAATCTCTTTCTCACCCTGGAGAGGAGGAGGAGG 2018
20 uAspArgProLeuGlyGlyGlyHis.....GlnProA 33
2019 GGACCGGGGACCGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2068
33 laglyAsn ArgArgGlyGlnAlaArgGlnLeuAlaProAsnPheArgTr 49

```



```

STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30B
CURRENT APPLICATION DATE: 07/95/1715A
FILING DATE: 25-SEP-1992
CLASSIFICATION: 800
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/772,027
FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Spitulnik, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: S-18805/A/COC 1577/CIP
TELEPHONE: (919)551-8615
TELEFAX: (919)551-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3624 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
FEATURES:
NAME/KEY: CDS
LOCATION: 1..3621
OTHER INFORMATION: /product= "Full-length, maize
OTHER INFORMATION: optimized cry1B"
OTHER INFORMATION: /note= "disclosed in Figure 6."
US-07-951-715A-6

alignment_scores:
Quality: 75.50 Length: 147
Ratio: 1.198 Gaps: 7
Percent Similarity: 42.857 Percent Identity: 25.850

alignment_block:
US-07-951-715A-6
Align seq 1/1 to: US-07-951-715A-6 from: 1 to: 3624
5 HsAGlGluAnAGlGluMetGluGluProMeGluAnAGlGluGluAs 21
2121 CACCACGAGGACGAGCAACTCTCCAG...CATCCACGAGGACGAGC 217
2121 PARGProLeuGlyGlyGlyGlyGlyHIsleGluProAlaGlyAsnARG 38
2121 GCGCTGTGTGGGGACGAGCAACATCACCATCCAGGAGGACGAGCGTG 221
2168 ArgGlnAla.....ARGargLeuAlaPro 45
2168 TTCACGAGCAACTACGTGACCTCCGCGACCTTCACAGAGGTGTCACC 2267
46 AsnPhe.....ArgTPal 50
2168 CACCTACTCTACCAAGAACTCGCGACGAGGACGTGAAGGCCGTACACCC 2317
50 alleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAsp... 65
2168 GCTATCCACTGCGCGGCTACATCTCAGGAGCACGACGAGCTGGAGATCTAC 2367
66 .....AspMet 67
2168 CTGATCCCTACAGCTTAAGCAGAGACCTCGACGCTGCTGCTGACCA 2417

```

```

APPLICANT: Merlin, Ellis J.
APPLICANT: Launis, Karen L.
APPLICANT: Rothstein, Steven J.
APPLICANT: Bowman, Cindy G.
APPLICANT: Dawson, John L.
APPLICANT: Runder, Gary M.
APPLICANT: Pace, Janet L.
APPLICANT: Suttie, Janet L.
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 601810artis Corporation
STREET: Patent & Trademark Dept., 520 White Plains
CITY: Tarrytown
STATE: New York
ZIP: 10591-9005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,595A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 800
PUBLICATION DATA: US 07/951,715
PUBLICATION NUMBER: 07/951,715
FILING DATE: 25-SEP-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/772,027
FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40403
REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)341-8582
TELEFAX: (919)341-8889
INFORMATION FOR SEQUENCE: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3624 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE KEY: CDS 3621
LOCATION: 1
OTHER INFORMATION: /product= "Full-length, maize
OTHER INFORMATION: optimized cryiB."
OTHER INFORMATION: /note= "disclosed in Figure 6."
US-08-459-595A-6
alignment_scores:
Quality: 75.50 Length: 147
Ratio: 1.198 Gaps: 7
Percent Similarity: 42.857 Percent Identity: 25.850
alignment_block:
US-09-327-750d-13 x US-08-459-595A-6 ..
Align seg 1/1 to: US-08-459-595A-6 from: 1 to: 3624
5 HlaGlnGluangluGluMetGluInProMetGlnAsnGluGluAs 21
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2121 CACCACGAGCAGACGACCTTCCACG...CATCCACGAGCAGC 2167

```

```

21 PATGPTLeuGlyGlyGluGlyHisGlnProMetGlnAsnGluGlu 38
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2168 ACAGCTGCTGGCGCAGCAGACACATCCACCATCCAGAGCGCACGCTG 2217
38 lyGlnAla.....ArgArgLeuAlaPro 45
|||||:|||||:|||||:|||||:|||||:|||||:
2218 TTCAGGAGAACTACGTGACCTGGCCGGGACCTTCACGAGTGCTACCC 2267
46 AsnPhe.....ArgTTrpAl 50
|||||:|||||:|||||:|||||:|||||:|||||:
2268 CACTACTCTCTACCAAGAGATCGCGGAGAGCCGAGCTGAGGCTACACCC 2317
50 alleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAsp..... 65
|||||:|||||:|||||:|||||:|||||:|||||:
2318 GCTACCACTCTGGCGGCTACATCGAGCAGACGAGGACCTGGAGATCTAC 2367
66 .....AapMet 67
|||||:|||||:|||||:|||||:|||||:|||||:
2368 CTGATCGCTTACACCCACGAGCAGGACCTCGGAGCTGCCCGGCGACCGA 2417
68 GluIlePheMetGluIleMetArgGluIleArgArgLysLeuArgGluLe 84
|||||:|||||:|||||:|||||:|||||:|||||:
2418 GACCTCTGGCCCTGTAGCGTGGAGAGCCGCCATCGCCCTCGCGGAGCC 2467
84 uGln...LeuArgAsnCysLeuArgIle..... 92
|||||:|||||:|||||:|||||:|||||:|||||:
2468 CCAACCGCTGGCGGCCCCACCTTCGAGTGGAGCCCGGACCTGGAGTGCAGC 2517
93 ..LeuMet.GlyGluLeuSerAsnHisHisAspHisHis 104
|||||:|||||:|||||:|||||:|||||:|||||:
2518 TCCCGCAGCGCGGAGAGTGGCGGCCACCCACGAGCCACAC 2556
seq_name: /cgn2.5/prodata/2/lna/6A_COMB.seq:US-08-459-504B-6
seq_documentation_block:
Sequence 6: Application US/08459504B
Patent No. 6075185
GENERAL INFORMATION:
APPLICANT: Kosziel, Michael G.
APPLICANT: Desai, Nalin M.
APPLICANT: Lewis, Kelly S.
APPLICANT: Kramer, Vance C.
APPLICANT: Warren, Gregory W.
APPLICANT: Evola, Stephen V.
APPLICANT: Crossland, Lyle D.
APPLICANT: Wierzbicki, Kristina S.
APPLICANT: Meriin, Ellis J.
APPLICANT: Launis, Karen L.
APPLICANT: Rothstein, Steven J.
APPLICANT: Bowman, Cindy G.
APPLICANT: Dawson, John L.
APPLICANT: Pace, Gary M.
APPLICANT: Dunder, Erik M.
APPLICANT: Suttie, Janet L.
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6075185artis Corporation
STREET: 3054 Cornwalis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: IBM PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,504B
FILING DATE:
CLASSIFICATION:

```


PR 07-JUN-1999; 99US-0327750.
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 XX Sato T;
 XX WPI; 2001-061707/07.
 XX New p75-neurotrophin receptor-associated cell death executor (NADE) and
 PT the gene encoding NADE, useful for modulating the activity of p75NTR
 PT and for detecting neurodegenerative diseases.
 XX
 XX Claim 12; Fig 1; 134pp; English.
 XX
 CC The present invention relates to a purified polypeptide capable of
 CC binding and modulating the activity of p75NTR. The polypeptide is
 CC binding and modulating the activity of p75NTR. The polypeptide is
 CC NGF-induced apoptosis, which plays an important role in neurogenetic
 CC diseases. The peptide of the invention and p75NTR are useful for
 CC inhibiting NF-kappaB activation in a cell or a subject, for inducing
 CC caspase-2 and caspase-3 activity to cleave poly (ADP-ribose) polymerase
 CC and fragment nuclear DNA in a cell by co-expression of (1) and p75-NTR.
 XX
 XX Sequence 700 BP; 177 A; 188 C; 203 G; 132 T; 0 other;

alignment_scores:
 Quality: 564.50 Length: 124
 Ratio: 5.086 Gaps: 1
 Percent Similarity: 89.516 Percent Identity: 82.258

alignment_block:

US-09-327-750D-13 x AAF23528

Align seg 1/1 to: AAF23528 from: 1 to: 700

1 MetalaAnIleHISGInGluAnGluMetGluGlnProMetGlnAs 17
 |||||
 177 ATGGCAATGTCACCAAGAAACAGAGAGCTGGAGAGCCCTGCAGAA 226
 |||||
 17 nGlyGluLusAspArgProLeuGlyGlyGlyGlyHISGInProAlaG 34
 |||||
 227 TGGACGAGAGAGCCCTGTGGAGAGAGTGGAGCCACCACTCTCTG 276
 |||||
 34 lyAsn.....ArgArg 37
 |||||
 277 CAACACACACACACACACACACACACACACACACACACACAG 326
 |||||
 38 GlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnAr 54
 |||||
 327 GGCAGGCTGGCCGACTTGGCCCTAACTTCGATGGGCGCATTCACACAG 376
 |||||
 54 gGlnIleAsnAspGlyMetGlyGlyAspGlyAspMetGluIlePheM 71
 |||||
 377 GCGATGATATCGGCTTGGGTGGAGATGGAGATGATATGAAATGTTC 426
 |||||
 71 etGluGluMetArgGluIleArgArgGlyLeuArgGluGluGlnLeuArg 87
 |||||
 427 TGGAGGAGATGAGAGAGATCCGAGAAAGCTTAGGGAGCTACAGCTGAG 476
 |||||
 88 AsnCysLeuArgIleLeuMetGlyGlyLeuSerAsnHisHisAspHisH 104
 |||||
 477 AATGTCTACGATCTTATGGGGGAGCTGCTTAACACACACAGATCACCA 526
 |||||
 104 aAspGluPheCysLeuMetPro 111
 |||||
 527 TGTATGATTCGCTTATGCT 548

seq_name: /SIDS7/scgdata/geneseq/geneseq/NA2000.DAT: AAC01005

seq_documentation_block:

ID AAC01005 standard; cDNA; 532 BP.
 XX

AAC01005;
 AC 06-OCT-2000 (first entry)
 XX Human secreted protein 5' EST, SEQ ID NO: 1003.
 XX Human; 5' EST, expressed sequence tag; secreted protein; cDNA isolation;
 XX gene therapy; chromosome mapping; ss.
 XX Homo sapiens.
 XX EPI033401-A2.
 XX 06-SEP-2000.
 XX 21-FEB-2000; 2000EP-0200610.
 XX 26-FEB-1999; 99US-0122487.
 XX (GEST) GENSET.
 XX Dumas Milne Edwards J, Duclert A, Giordano J;
 XX WPI; 2000-500381/45.
 XX P-PSDB; AAG00999.
 XX
 CC New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 CC obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 CC diagnostic, forensic, gene therapy and chromosome mapping procedures -
 CC Claim 1; SEQ ID 1003; 71pp + CD-ROM; English.
 XX
 CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.
 XX
 XX Sequence 532 BP; 151 A; 121 C; 163 G; 91 T; 6 other;

alignment_scores:

Quality: 537.00 Length: 104
 Ratio: 5.424 Gaps: 0
 Percent Similarity: 95.192 Percent Identity: 95.192

alignment_block:

US-09-327-750D-13 x AAC01005

Align seg 1/1 to: AAC01005 from: 1 to: 532

1 MetalaAnIleHISGInGluAnGluMetGluGlnProMetGlnAs 17
 |||||
 219 ATGGCAATATTCACCGAGAAACAGAGAGATGGAGAGCTATGCAGAA 268
 |||||
 17 nGlyGluLusAspArgProLeuGlyGlyGlyGlyHISGInProAlaG 34
 |||||
 269 TGGAGGAGAGAGCCCTTGGAGAGAGTGGAGCCACCACTGCAG 318
 |||||
 34 lyAsnArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgProAla 50
 |||||
 319 GAATTCGACGGGAGAGCTCGCCGAGTTCGCCCTAATTTTCGATGGGCC 368
 |||||
 51 IleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspMet 67

Claim 5; Page 181-182; 205pp; English.

Sequences AAP59590-BAF59643 represent cDNAs encoding 54 human cell cycle and proliferation proteins (CCVPR), AAB60453-CCVPR and agonists of CCVPR are used to treat diseases or conditions associated with decreased expression of functional CCVPR, while CCVPR antagonists are used to treat diseases or conditions associated with overexpression of functional CCVPR. Monoclonal or polyclonal antibodies to CCVPR may be used in enzyme-linked immunosorbent assays (ELISA) or radioimmunoassays to detect CCVPR. CCVPR itself may be used to detect compounds (e.g., antibodies, oligonucleotides and proteins) that interact with CCVPR. CCVPR antagonists and agonists may be used to identify compounds that modulate the activity of CCVPR. CCVPR nucleotides can be used to generate transgenic animal models of human disease, and can be used in gene therapy in target cells with genetic abnormalities with respect to the expression of CCVPR for the treatment or prevention of a disorder associated with CCVPR. Diseases which can be diagnosed, treated and prevented using CCVPR proteins, nucleic acids, agonists or antagonists include immune, developmental and cell signalling disorders, and cell proliferative disorders including cancer. Specific examples of these disorders include Down's syndrome, epidermolysis, ataxia-telangiectasia, immunodeficiencies, diabetes mellitus, disorders of the menstrual cycle and infections caused by bacteria.

[illegible]


```

US-09-327-750D-13 x AAC10889
...
Align seq 1/1 to: AAC10889 from: 1 to: 692

3 AsnIleHisGluGluGluGluGluMetGluGlnProMetGlnAsnGlyGI 19
282 ANTCTCACACGAAATATGAAAGAAGATGAAGAAGCAGCACTGCTAA 331
19 uGluuGargProGluGly.....GlyGluGluGlyHisGlnP 32
332 TAAGGAGGAGCCCTTGCCCTACCTTTGGATGTGTGCAATCATGTGTC 381
32 -oALeGluVAsnArgArgGly...GlnAlaArgArgLeuIaProAsnPhe 47
382 CTAGAGGAATCTGTAGCGGTTGCGCCTTAGCAGCACCCTCTGCAGTAT 431
48 ArgTrpAlaIleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGI 64
432 AGCTGGGATGTATGCTATAGCTTGGAGAGACACAGCGCAGGATGANA 481
64 VAspAspMetGluIlePheMetGluGluMetArgGluIleArgGlySL 81
482 AGAGAATATGGAAGGATTTGGGGAGGAGGTAGACAGCTGATGGAAAGC 531
81 euArgGluGluGluGluArgGlyGlyGluGluGluGluGluGluGlu 97
532 TGGAGAAAGAGATTTGCTAGCTCTCGGCATCATGACGATGAC... 578
98 SerAsnHisHisAspHisHisAspClyPheCysLeuMetPro 111
579 CCCCCTCCACCATGACCATCATGATGAGCTTTTGCNNNATGCCC 620

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sv model

Run on: March 11, 2002, 15:42:51 ; Search time 144 Seconds
(Without alignments)
4167.553 Million cell updates/sec

Title: US-09-327-750d-28
Perfect score: 700
Sequence: 1 acagcgtctgcccagcagc.....ggggcttgggtccagcga 700

Scoring table: IDENTITY_NUC
Gapop 10.0, Capext 1.0

Searched: 930621 seqs, 42862619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_1101.1
1: /SID22/gcgdata/geneseq/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/NA2001.DAT.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	693.6	99.1	700	22	AAF23528
2	307.4	43.9	891	22	AAF23529
3	284.6	40.7	891	22	AAF23530
4	284.6	40.7	891	22	AAF23531
5	102.2	27.5	485	21	AAO1004
6	147.2	21.0	187	21	AAH7147
7	127.2	18.2	426	22	AAH2586
8	80.2	11.5	865	22	AAH0317
9	80.2	11.5	1229	22	AAH13750
10	80.2	11.5	1364	22	AAH5548
11	58	8.3	401	21	AAO1003

12	56.4	8.1	792	22	AAH45143
13	55.8	8.0	767	22	AAH75810
14	55.8	8.0	898	22	AAH75811
15	52.6	7.9	892	21	AAO1006
16	52.6	7.9	892	21	AAO1007
17	55.6	7.9	698	21	AAO1008
18	55.6	7.9	698	21	AAO1009
19	54.2	7.7	858	22	AAH50367
20	50.4	7.2	862	22	AAH50368
21	44.6	6.4	5552	21	AAH58182
22	42.2	6.0	128	21	AAH58183
23	41.4	5.9	830	22	AAH58184
24	41.4	5.9	830	22	AAH58185
25	41.4	5.9	830	22	AAH58186
26	40.8	5.8	381	22	AAH58187
27	40.8	5.8	381	22	AAH58188
28	40.8	5.8	381	22	AAH58189
29	40.8	5.8	381	22	AAH58190
30	40.4	5.8	10732	21	AAH10594
31	39.2	5.6	1635	22	AAH12780
32	39.2	5.6	1635	22	AAH12781
33	39.2	5.6	1635	22	AAH12782
34	39.2	5.6	1635	22	AAH12783
35	39.2	5.6	1635	22	AAH12784
36	39.2	5.6	1635	22	AAH12785
37	38.8	5.5	1452	21	AAH26304
38	38.6	5.5	1452	21	AAH26305
39	38.4	5.5	1185	21	AAH26306
40	38.4	5.5	1185	21	AAH26307
41	38.4	5.5	1185	21	AAH26308
42	37.6	5.4	439	22	AAH26309
43	37.6	5.4	439	22	AAH26310
44	37.6	5.4	439	22	AAH26311
45	37.6	5.4	439	22	AAH26312

ALIGNMENTS

RESULT 1
AAF23528
AAF23528 standard; DNA; 700 BP.
AAF23528:
AC AAF23528:
DT 22-MAR-2001 (first entry)
DE Mouse NADE DNA.
KW Neurotrophin receptor; p75-NTR; NGF-induced apoptosis;
KW Neurogenetic disease; NF-kappaB; ds.
KW Mus sp.
XX WO200075278-A2.
XX 14-DEC-2000.
XX 07-JUN-2000; 2000WO-US15621.
XX 07-JUN-1999; 99US-0327750.
XX (UNCO) UNIV COLUMBIA NEW YORK.
XX Sato T.
XX WPI; 2001-061707/07.
XX New p75-neurotrophin receptor-associated cell death executor (NADE) and
XX the gene encoding NADE, useful for modulating the activity of p75NTR
XX and for detecting neurodegenerative diseases -
XX Claim 12; Fig 1; 134pp; English.

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 11, 2002, 14:17:05 ; Search time 1319.57 Seconds
(without alignments)
8751.355 Million cell updates/sec

Title: US-09-327-750d-28

Perfect score: 700

Sequence: 1 acagcgtctgcgcagcagc.....ggggcttgcgttcagctga 700

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank:

1: gb:ba.*

2: gb:htg.*

3: gb:in.*

4: gb:om.*

5: gb:ov.*

6: gb:pat.*

7: gb:ph.*

8: gb:pl.*

9: gb:pr.*

10: gb:ro.*

11: gb:sts.*

12: gb:sv.*

13: gb:un.*

14: gb:va.*

15: gb:ba.*

16: em:fun.*

17: em:hum.*

18: em:in.*

19: em:om.*

20: em:or.*

21: em:ov.*

22: em:pat.*

23: em:ph.*

24: em:pl.*

25: em:sta.*

26: em:sv.*

27: em:un.*

28: em:va.*

29: em:htgo_hum.*

30: em:htgo_inv.*

31: em:htgo_inv.*

32: em:htgo_inv.*

33: em:htgo_inv.*

34: em:htgo_inv.*

35: em:htgo_inv.*

36: em:htgo_inv.*

37: em:htgo_inv.*

38: em:htgo_inv.*

39: em:htgo_inv.*

40: em:htgo_inv.*

41: em:htgo_inv.*

42: em:htgo_inv.*

43: em:htgo_inv.*

44: em:htgo_inv.*

45: em:htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

J: Biol. Chem. 275 (23), 17566-17570 (2000)

20298829

Result No.	Score	Query Match	Length	DB ID	Description
1	62.2	98.2	700	10 AF187066	AF187066 Mus muscu
2	62.2	98.2	700	10 AF187066	AF187066 Mus muscu
3	35.8	50.8	519	10 AF187065	AF187065 Rattus nor
4	33.8	48.3	733	9 BC003190	BC003190 Homo sapi
5	307.4	43.9	891	9 HM000000	M31888 Human unkno
6	307.4	43.9	891	9 AF187064	AF187064 Homo sapi
7	240.6	34.4	421	11 G72708	G72708 MARC 4553-4
8	154.8	22.1	98274	9 AF097439	AF097439 Mus muscu
9	154.8	22.1	98274	9 HS71487	HS71487 Human DNA s
10	101.4	14.5	477	11 G35294	G35294 Human DNA s
11	101.4	14.5	477	11 G35294	G35294 Human DNA s
12	80.2	11.5	1229	9 AK000959	AK000959 Homo sapi
13	80.2	11.5	1229	9 AK000959	AK000959 Homo sapi
14	79.2	11.1	6948	6 HS035649	HS035649 Human DNA
15	67.6	9.7	59	10 M5000000	M5000000 Mus muscu
16	67.6	9.7	59	10 M5000000	M5000000 Mus muscu
17	59.8	8.5	7218	6 AF183416	AF183416 Homo sapi
18	59.8	8.5	7218	6 AF183416	AF183416 Homo sapi
19	59.8	8.5	7218	6 AF183416	AF183416 Homo sapi
20	59.8	8.5	7218	6 AF183416	AF183416 Homo sapi
21	57.2	8.2	43952	9 HSU10564	HSU10564 Human DNA s
22	56.4	8.1	791	9 AF237783	AF237783 Homo sapi
23	55.8	8.0	744	9 AF251053	AF251053 Homo sapi
24	55.8	8.0	744	9 AF251053	AF251053 Homo sapi
25	55.8	8.0	744	9 AF251053	AF251053 Homo sapi
26	55.8	8.0	744	9 AF251053	AF251053 Homo sapi
27	55.8	8.0	744	9 AF251053	AF251053 Homo sapi
28	54.2	7.7	4058	11 G72708	G72708 MARC 4553-4
29	54.2	7.7	4058	11 G72708	G72708 MARC 4553-4
30	54.2	7.7	4058	11 G72708	G72708 MARC 4553-4
31	54.2	7.7	4058	11 G72708	G72708 MARC 4553-4
32	49.2	7.0	185306	2 AC013570	AC013570 Homo sapi
33	48.8	7.0	37224	3 AC007061	AC007061 Leishmani
34	48.2	6.9	36823	9 HSU20961	HSU20961 Human DNA s
35	46	6.6	14577	2 AL357046	AL357046 Homo sapi
36	46	6.6	14577	2 AL357046	AL357046 Homo sapi
37	46	6.6	14577	2 AL357046	AL357046 Homo sapi
38	44.8	6.4	190544	9 AC005768	AC005768 Homo sapi
39	44.2	6.3	110000	2 LMP1CHR15_5	Continuation (6 of
40	44.2	6.3	172200	2 AC022626	AC022626 Homo sapi
41	43.2	6.2	210211	3 AC070038	AC070038 Dictyoste
42	42.8	6.1	164025	2 PRCGSPR	PRCGSPR Homo sapi
43	42.6	6.1	164025	2 PRCGSPR	PRCGSPR Homo sapi
44	42.6	6.1	164025	2 PRCGSPR	PRCGSPR Homo sapi
45	42.6	6.1	164025	2 PRCGSPR	PRCGSPR Homo sapi

ALIGNMENTS

RESULT 1
AF187066 700 bp mRNA 11-JUN-2000
Mus musculus p75NTR-associated cell death executor (Nade) mRNA,
Complete cds.
ACCESSION AF187066
VERSION AF187066.1 GI:8452897
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 700)
AUTHORS Weir, J. J., Hoshino, S., Kimura, M. T., Nadano, D.,
Suzuki, P., Hoshino, S., Hoshino, S., and Saito, T. A.
TITLE Nade, a p75NTR-associated cell death executor, is involved in
signal transduction mediated by the common neurotrophin receptor
p75NTR
J: Biol. Chem. 275 (23), 17566-17570 (2000)
MEDLINE 20298829

[illegible]


```

Tel: 4157259687
Fax: 4157259689
Email: myers@hgc.stanford.edu

Primer A: ACATCTTTCCTCAGCAAGTTGATC
Primer B: TTCTTGCACTCTCTGCA
STS ID: 106
PCR Profile:
  Initial incubation: 95 degrees C for 10 minutes
  Denaturation: 94 degrees C for 15 seconds
  Annealing: 60 degrees C for 30 seconds
  Polymerization: 72 degrees C for 23 seconds
  PCR Cycles: 30
  Thermal Cycler: Perkin Elmer 9600

Protocol:
  Template: 25 ng
  Primer: each 1 uM
  dNTPs: each 200 uM
  Amplifrag Gold Polymerase: 5 uL
  Total Vol: 10.07 units/uL

Buffer:
  MgCl2: 2.5 mM
  KCl: 50 mM
  Tris-HCl: 10 mM
  pH: 8.3

Prepared with primer pairs provided by Sandoz, derived from NS1315
-- Washington University/Merck EST sequence.
FEATURES
  source
    1..477
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    STS
    primer_bind 44..67
    primer_bind complement(130..149)
  BASE COUNT 153 a 115 c 79 g 130 t
  ORIGIN
    Query Match 14.5%; Score 101.4; DB 11; Length 477;
    Best Local Similarity 72.9%; Prev. No. 4,5e-16;
    Matches 161; Conservative 0; Mismatches 31; Indels 9; Gaps 2;
    Oy 339 cgactgcgcccttaactccgattccgattcccaacagcc-----agctgaatgacggg 392
    Db 438 CCACCTGCGCCGGTTTGGAGAGGATGTGCCAATAGCTGTGCGATACATTGATG 379
    Oy 393 ttgggtgagatggagatgatataatgttcatggagagatgagagatccggaga 452
    Db 378 ATAGATGAGATGGAGATGATGATGAAACGGTTTCATGGAGGAGATGAGAGCTTAAGAGG 319
    Oy 453 aagcttagagagctacgtcgagaaatgtctacgcatctcttctgctgcttctaac 512
    Db 318 AAMATTAGGAACTTCATGTGAGGTACAGTCTCCCATTTCTTTAGGGGACC---CTCT 262
    Oy 513 caccacgatcaccatgatgaattctcttctcttctgcttctgact 553
    Db 261 CACCATGATCATCATGATGAGTTTGCCTTATGCCCTTGAAT 221

RESULT 11
NS1315      HSV351F8  45678 bp  DNA  PRI  23-NOV-1999
LOCUS      Human DNA sequence from cosmid V351F8, between markers DXS366 and
DEFINITION DXS87 on chromosome X contains ESTs.
ACCESSION  270719
VERSION    270719.1 GI:1261915
KEYWORDS   X.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (Bases 1 to 45678)
AUTHORS   Whiteley, M.
TITLE     Direct Submission
JOURNAL   Submitted (09-APR-1995) Sanger Centre, Hinxton, Cambridgeshire,
          CB10 1RQ, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT   IMPORTANT: This sequence is the entire insert of clone V351F8. The
          true left end of clone V351F8 is at 1 in this sequence. The true
          right end of clone V351F8 is at 45678.
          V351F8 is from the human chromosome X-specific cosmid library.
FEATURES   location/Qualifiers
  source
    1..45678
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="X"
    /map="X"
    /clone="GHC-351F8"
    /clone_lib="SCCV"
    1..70
  repeat_region
    /partial
    /note="Alu repeat: matches 80..1 of consensus"
    642..930
  repeat_region
    /note="Alu repeat: matches 1..308 of consensus"
    1468..5179
  repeat_region
    /note="L1 element fragment"
    5243..5276
  repeat_region
    /note="17 copies of 2 mer 82 & conserved"
    5279..5567
  repeat_region
    /partial
    /note="Alu repeat: matches 304..1 of consensus"
    9975..10369
  repeat_region
    /note="MSTA element fragment"
    10399..12017
  repeat_region
    /note="MSTAR element fragment"
    1318..11576
  repeat_region
    /note="MSTAR element fragment"
    11715..11846
  repeat_region
    /note="MTR8 element fragment"
    12074..12170
  repeat_region
    /note="MSTA element fragment"
    12227..12486
  repeat_region
    /note="MSTA element fragment"
    12437..12502
  repeat_region
    /note="MSTC element fragment"
    12756..13059
  repeat_region
    /partial
    /note="Alu repeat: matches 308..1 of consensus"
    14478..14650
  repeat_region
    /note="THE1B element fragment"
    14491..14536
  repeat_region
    /note="MSTA element fragment"
    14715..14831
  repeat_region
    /note="THE1B element fragment"
    15099..15533
  repeat_region
    /note="L1 element fragment"
    15534..15626
  repeat_region
    /note="31 copies of 3 mer 85 & conserved"
    15868..15819
  repeat_region
    /note="31 copies of 18 mer 98 & conserved"
    16352..16436
  repeat_region
    /note="L1 element fragment"
    16728..16763
  repeat_region
    /note="9 copies of 4 mer 94 & conserved"
    16764..17054
  repeat_region
    /partial
    /note="Alu repeat: matches 308..1 of consensus"
    19420..19449
  repeat_region
    /note="15 copies of 2 mer 87 & conserved"
    complement(19537..20013)
  misc_feature
    complement(19537..20013)
  repeat_region
    /note="match: 3' EST NS1315 clone 283069"
    20071..20114
  repeat_region
    /note="22 copies of 2 mer 98 & conserved"
    20073..20112

```


repeat_region 27976..28273

Query Match 11.54; Score 80.2; DB 9; Length 69648;
Best Local Similarity 56.24; Pred. No. 2.4e-10;
Matches 269; Conservative 0; Mismatches 153; Indels 57; Gaps 4;

QY 143 gccagatgacacccggagacacaaatctcatctatgccaatgtccaccagaaacga 202
DB 64298 GTCCAGAGAGAACTACCGGCAACAAATCTCACGGGAAATCCCAACAGAAACGA 64357
QY 203 agagctggagcag---ccctcgagagagagacacacccctctgaggagaggtga 259
DB 64358 AGGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 64417
QY 260 gggccaccagctgctgcacacacacacacacacacacacacacacacacacac 319
DB 64418 AGGCCAGAGAGCTCGAGAGAAATATCA----- 64443
QY 320 ccagagagcagcagctgagcagcagcagcagcagcagcagcagcagcagcagc 379
DB 64444 -GGCGGGCGAGGTAGGCGAGTCTCTTAATTTTCGTGGGCCACCTAATAGCA 64501
QY 380 gatgaatgagcgggttggggagagatgagatgagatgagatgagatgagatgag 439
DB 64502 TATTGAGCAAA-----TGAAGCGAGAGATGATGAGAAAGGTTGTAGGCGAGATGAT 64555
QY 440 agagatccggagagctaggagcagcagcagcagcagcagcagcagcagcagcag 499
DB 64556 GGAATATCAAGAGAAAGACTAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 64611
QY 500 gggagctctcaac 559
DB 64612 -----CCTGAGCTGACACACCTATGACTTTGCTCTACCTTACCTTACCTTAA 64663
QY 560 attcccccagagac 618
DB 64664 AGTTTCCTGAGGTAAATGAGACACTGCTTACAGCTGATTTTGTGATTTACT 64722

RESULT 15
LOCUS G13550 244 bp DNA STS 04-JAN-1996
DEFINITION Human STS SHGC-11140.
ACCESSION G13550
VERSION G13550.1 GI:1129289
KEYWORDS STS sequence; primer; sequence tagged site.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Hominidae; Homo.
REFERENCE Myers, R.M.
AUTHORS Myers, R.M.
JOURNAL Unpublished (1995)
COMMENT Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, H-344, Stanford, CA 94305, USA
Tel: 415/7259687
Fax: 415/7259689
Email: myers@shgc.stanford.edu
Primer A: CCCTTGTGCTGAACCTAT
Primer B: TTAAATTTTCCATGCAATGG
STS size: 138
PCR Profile:
Initial incubation: 94 degrees C for 90 seconds
Denaturation: 94 degrees C for 15 seconds
Annealing: 52 degrees C for 30 seconds
Polymerization: 72 degrees C for 30 seconds

PCR Cycles: 30 Perkin Elmer 9600
Thermal Cycler: 25 ng each 1 uM
Template: each 200 uM
dNTPs: each 200 uM
Taq Polymerase: 0.05 units/uL
Total Vol: 10 uL
Buffer: MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 20 mM
pH: 8.3
Prepared with primer pairs derived from M38188 -- Unigene.
Location/Qualifiers
L: 244
Organism: "Homo sapiens"
STS
Primer_bind 84..103
Primer_bind complement(199..221)
BASE COUNT 57 a 50 c 42 g 95 t
ORIGIN
Query Match 11.14; Score 77.4; DB 11; Length 244;
Best Local Similarity 80.24; Pred. No. 8.6e-10;
Matches 105; Conservative 0; Mismatches 21; Indels 5; Gaps 1;
QY 567 cctgagac 626
DB 15 CATGATTAATACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 74
QY 627 tttaataac 686
DB 75 TTTACTGATCCGTTTGTGTCGACCTATGTTT-----CCATGTCGCAAGTGGTCT 129
QY 687 tgtgtgtgcag 697
DB 130 TGTGTTGGCAG 140
Search completed: March 11, 2002, 16:33:03
Job time: 8136 sec

④


```
|||||
Db 1 CCCGGGGGCTGGCAGCGGGCCCGAGCGGAGCGGACAGATTGACTGGAAGCCGAGATC 60
QY 97 caggcgagcggggaattgacaggaggactacgcccgaagggtatggcccaggaatagcaa 156
|||||
Db 61 CAGCGCGGACCGGGAAATTGACAGGAGACTACCGCCGCAAGGATAGGCCCCAGATAGCAA 120
QY 157 cagggaacaacatctcatcgccaatctccaccagggaacaggaagactggagcagc 216
|||||
Db 121 CAGGAACAACAAATCTCATCATGTGCCCATGTCCACGAGGAACGAAGATGGAGCAGC 180
QY 217 ccctgcagaatggacaggaaacaccgcccctgtggagaggtgagggccaccagcctgctg 276
|||||
Db 181 CCCTGCGAATGGACAGGAAGACCGCCCTGTGGGAGAGGTGAGGGCCACCCAGCCTGCTG 240
QY 277 caaicaacaacaacaacaacaacaacaacaacaacaacaacaacaacaacccgagaggctc 336
|||||
Db 241 CAACACACACACACACACACACACACACACACACACACACACACACACACACACAC 300
QY 337 gccgcacctgcaccttaacttcctcgatggggcattcccaacaggcagatgacgggtgg 396
|||||
Db 301 GCCGACTTGCCCTAACTTCCGATGGGCCATTCCCACAGGCGAGATGAATGACGGGTTGG 360
QY 397 gtggagatggagatgatggaaatgttcctatggaggagatgagagagatccggagaagc 456
|||||
Db 361 GTGGAGATGGAGATCATGGGAATTTTCATGGAGGAGATGAGAGATCCGGAGAAAGC 420
QY 457 ttagggactacagctgagaatgtctacgacatccttatgggggagctgtctaaacac 516
|||||
Db 421 TTAGGGACTACAGCTGAGAAATTTGCTACGCATCCTTATGGGGAGCTGTCTACACACC 480
QY 517 acgatcacatgatgaattctgcttactgcttactgcttactgcttactcccccctgagatcc 576
|||||
Db 481 ACGATCACCATGATGAATCTGCCTTATGCCTTGACTTCGGTCAATTCG-CCCTGAGATCC 539
QY 577 atactgtactcc 589
|||||
Db 540 ATACTGTGACTCC 552
```

Search completed: March 11, 2002, 16:09:50
Job time: 10885 sec

APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOXP2 VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
STREET: 1801 O'Connell Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: US/08/232,463
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
COUNTRY OF ORIGIN: 0472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 1836-9300
TELEFAX: (703) 1683-4109
TELEX: 895149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
INVERSE SOURCE:
CLONE: T79pt-f18
US-08-232-463-14

Query Match 5.5%; Score 38.4; DB 1; Length 7218;
Best Local Similarity 7.6%; Pred. No. 0.12;
Matches 18; Conservative 126; Mismatches 92; Indels 0; Gaps 0;
OY 441 gagatccgggaagcttagggagctacagctgagaaattgtctacgcatcctattgagg 500
DB 1035 GAGTGTCTGCGAGGTCGAGGAGGCTGCGAGGCTGCGAGGCTGCGAGGCTGCGAGG 1094
OY 501 gacgtgtctaacaccac 560
DB 1095 YY 1154
OY 561 tcccccttgagatcattgtgactccgctgtagcccttcctccctcatttctcga 620
DB 1155 YY 1214
OY 621 catgcttaagaccggttctggtgagccctgtgtattctccatgcatgccc 676
DB 1215 YY 1270

RESULT 4
US-09-007-005-17
Sequence 17, Application US/09007005B
Patent No. 6258558
GENERAL INFORMATION:
APPLICANT: Szostak, Jack W.
APPLICANT: Roberts, Richard W.

APPLICANT: LIU, Rihe
TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
FILE REFERENCE: 00786/350003
CURRENT APPLICATION NUMBER: US/09/007,005B
EARLIER FILING DATE: 1998-01-14
EARLIER APPLICATION NUMBER: 60/035,963
EARLIER FILING DATE: 1997-11-06
EARLIER APPLICATION NUMBER: 60/064,491
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17
LENGTH: 289
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Translation template
NAME: r18c_feature
LOCATION: (1..289)
OTHER INFORMATION: n - A,T,C or G
US-09-007-005-17

Query Match 5.3%; Score 36.8; DB 4; Length 289;
Best Local Similarity 6.5%; Pred. No. 0.056;
Matches 15; Conservative 99; Mismatches 118; Indels 0; Gaps 0;
OY 98 agggcgccgggaattgacggagactacgcgcgaaggagatagcccgacgaatcgcaac 157
DB 1 tggggtcctatctatctatctatctatctatctatctatctatctatctatctatctat 60
OY 158 caggagacaaatctctatctatctatctatctatctatctatctatctatctatctat 217
DB 61 tctatctatctatctatctatctatctatctatctatctatctatctatctatctatct 120
OY 218 cctgagatgacgagac 277
DB 121 tctatctatctatctatctatctatctatctatctatctatctatctatctatctatct 180
OY 278 aac 329
DB 181 tctatctatctatctatctatctatctatctatctatctatctatctatctatctatct 232

RESULT 5
US-09-244-796-17
Sequence 17, Application US/09244796
Patent No. 6281344
GENERAL INFORMATION:
APPLICANT: Szostak, Jack W.
APPLICANT: Roberts, Richard W.
TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
FILE REFERENCE: 00786/350007
CURRENT APPLICATION NUMBER: US/09/244,796
EARLIER FILING DATE: 1999-02-05
EARLIER APPLICATION NUMBER: 60/035,963
EARLIER FILING DATE: 1997-01-27
EARLIER APPLICATION NUMBER: 60/064,491
EARLIER FILING DATE: 1997-11-06
EARLIER APPLICATION NUMBER: 09/007,005
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17
LENGTH: 289
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Translation template

```

1 TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS
2
3 TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
4
5 TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
6
7 TITLE OF INVENTION: SPECIES INFECTIONS
8
9 NUMBER OF SEQUENCES: 30
10
11 CORRESPONDENCE ADDRESS:
12
13 ADDRESSEE: PETERS, VERNY, JONES & BIKSA
14 STREET: 385 Sherman Avenue, Suite 6
15 CITY: Palo Alto
16 STATE: CA
17 COUNTRY: USA
18
19 ZIP: 94306-1840
20
21 COMPUTER READABLE FORM:
22 MEDIUM TYPE: Floppy disk
23
24 COMPUTER: IBM PC compatible
25
26 OPERATING SYSTEM: PC-DOS/MS-DOS
27
28 SOFTWARE: Patentin Release #1.0, Version #1.30
29
30 CURRENT APPLICATION DATA: /08/928, 361B
31
32 FILING DATE: 12-SEP-1997
33
34 CLASSIFICATION:
35
36 PRIOR APPLICATION DATA:
37
38 APPLICATION NUMBER: US 60/026,062
39
40 FILING DATE: 13-SEP-1996
41
42 ATTORNEY/AGENT INFORMATION:
43
44 NAME: Verny, Hana
45
46 REGISTRATION NUMBER: 30,518
47
48 REFERENCE/DOC#T NUMBER: 480,76-1(HV)
49
50 TELEPHONE: 650-324-1677
51
52 TELEFAX: 650-324-1678
53
54 INFORMATION FOR SEQ ID NO: 3:
55
56 SEQUENCE CHARACTERISTICS:
57 LENGTH: 5318 base pairs
58 TYPE: nucleic acid
59 STRANDEDNESS: double
60 TOPOLOGY: linear
61 MOLECULE TYPE: DNA (genomic)
62
63 US-08-928-361B-3
64
65 Query Match 5.0%; Score 34.8; DB 3; Length 5318;
66 Best Local Similarity 84.8%; Pred. No. 1,3;
67 Matches 39; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
68
69 QY 279 aaacaaacacacacacacacacacacacacacacacacacacacacacagaa 324
70 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
71 DB 873 AACTACACACACACACACACACACACACACACACACACACACACACAACTACCAAGNA 918
72
73 RESULT 11
74 US-08-692-922-1/G
75 Sequence 1, Application US/08692922
76 Patent No. 6277592
77 GENERAL INFORMATION:
78 APPLICANT: Bidwell, Christopher A.
79 APPLICANT: Spurluck, Michael E.
80 TITLE OF INVENTION: PORCINE LEFTIN PROTEIN, NUCLEIC ACID
81 TITLE OF INVENTION: SEQUENCES CODING THEREFOR AND USES THEREOF
82 NUMBER OF SEQUENCES: 30
83 CORRESPONDENCE ADDRESS:
84 ADDRESSEE: WHYTE HIRSCHBOECK DUDEK S.C.
85 STREET: Suite 2100 111 East Wisconsin Avenue
86 CITY: Milwaukee
87 STATE: Wisconsin
88 COUNTRY: USA
89 ZIP: 53202
90
91 COMPUTER READABLE FORM:
92 MEDIUM TYPE: Floppy disk
93 OPERATING SYSTEM: IBM PC compatible
94 SOFTWARE: Patentin Release #1.0, Version #1.30
95
96 CURRENT APPLICATION DATA:

```

Query Match 4.9%; Score 34.4; DB 4; Length 1247;
Best Local Similarity 59.0%; Pred. No. 0.72; Indels 0; Gaps 0;
Matches 59; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY 229 gacaggaacacccgctgtggagaggtgagggccacccagctgtctcaaacacacac 288
DB 1100 gactgtaccattgcactctggcctggcgacagcagcctgtctcaaaaagaaa 1041
QY 289 ac 328
DB 1040 acaaaaac 1001

RESULT 14
US-08-787-739-58/c 4.9%; Score 34.4; DB 4; Length 1247;
Patent No. 6027887
GENERAL INFORMATION:
APPLICANT: Pastorekova, Silvia
APPLICANT: Pastorek, Jaromir
TITLE OF INVENTION: MN Gene and Protein
NUMBER OF SEQUENCES: 96
CORRESPONDENCE ADDRESS:
ADDRESSEE: Leona L. Lauder
STREET: 369 Pine Street, Suite 610
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION NUMBER: US/08/787,739
FILING DATE: 24-JAN-1997
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: US 08/485,049
FILING DATE: 07-JUN-1995
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: US 08/486,756
FILING DATE: 07-JUN-1995
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: US 08/477,504
FILING DATE: 07-JUN-1995
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: US 08/481,658
FILING DATE: 07-JUN-1995
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: US 08/485,862
FILING DATE: 07-JUN-1995
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: US 08/485,863
FILING DATE: 07-JUN-1995
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: US 08/487,077
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Lauder, Leona L.
REGISTRATION NUMBER: 30,863
TELEPHONE: 415-981-2034
TELEFAX: 415-981-0332
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 2501 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
DESCRIPTION: MN genomic region between pMN1 and
DESCRIPTION: B03 assumed to contain regulatory
ELEMENTS
HYPOTHEtical: NO
ANTI-SENSE: NO
US-08-787-739-58
Query Match 4.9%; Score 34.4; DB 3; Length 2501;
Best Local Similarity 59.0%; Pred. No. 1.1;
Matches 59; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY 229 gacaggaacacccgctgtggagaggtgagggccacccagctgtctcaaacacacac 288
DB 2421 gactgtaccattgcactctggcctggcgacagcagcctgtctcaaaaagaaa 2362
QY 289 ac 328
DB 2361 acaaaaac 2322
RESULT 15
US-09-178-115-58/c 4.9%; Score 34.4; DB 4; Length 2501;
Sequence 58, Application US/09178115
Patent No. 6297610;
GENERAL INFORMATION:
APPLICANT: Pastorekova, Silvia
APPLICANT: Pastorek, Jaromir
TITLE OF INVENTION: MN Gene and Protein
FILE REFERENCE: D-0021.5A
CURRENT APPLICATION NUMBER: US/09/178,115
FILING DATE: 1998-10-23
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: 09/177,776
FILING DATE: 1998-10-23
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: 08/787,739
FILING DATE: 1997-01-24
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: 08/485,049
FILING DATE: 1995-06-07
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: 08/486,756
FILING DATE: 1995-06-07
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: 08/477,504
FILING DATE: 1995-06-07
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: 08/481,658
FILING DATE: 1995-06-07
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: 08/485,862
FILING DATE: 1995-06-07
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: 08/485,863
FILING DATE: 1995-06-07
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: 08/487,077
FILING DATE: 1995-06-07
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: 08/260,190
FILING DATE: 1994-06-15
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: 08/177,093
FILING DATE: 1993-12-30
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: 07/964,589
FILING DATE: 1992-10-21
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: PV-709-92
FILING DATE: 1992-03-11
NUMBER OF SEQ ID NOS: 116
SOFTWARE: Patent Ver. 2.0
SEQ ID NO: 58
DESCRIPTION: MN Gene and Protein
TYPE: DNA
ORGANISM: HUMAN
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(2501)
US-09-178-115-58
Query Match 4.9%; Score 34.4; DB 4; Length 2501;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sv model

Run on: March 11, 2002, 16:36:39 ; Search time 144 Seconds
(Without alignments)
5304.700 Million cell updates/sec

Title: US-09-327-750d-29

Perfect score: 891
Sequence: 1 accccatccccctctat.....ataaagcaatttaanaagc 891

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs. 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N_Geneseq_1101.*
- 1: /SIDS2/gcgdata/geneseq/NA1980.DAT.*
 - 2: /SIDS2/gcgdata/geneseq/NA1981.DAT.*
 - 3: /SIDS2/gcgdata/geneseq/NA1982.DAT.*
 - 4: /SIDS2/gcgdata/geneseq/NA1983.DAT.*
 - 5: /SIDS2/gcgdata/geneseq/NA1984.DAT.*
 - 6: /SIDS2/gcgdata/geneseq/NA1985.DAT.*
 - 7: /SIDS2/gcgdata/geneseq/NA1986.DAT.*
 - 8: /SIDS2/gcgdata/geneseq/NA1987.DAT.*
 - 9: /SIDS2/gcgdata/geneseq/NA1988.DAT.*
 - 10: /SIDS2/gcgdata/geneseq/NA1989.DAT.*
 - 11: /SIDS2/gcgdata/geneseq/NA1990.DAT.*
 - 12: /SIDS2/gcgdata/geneseq/NA1991.DAT.*
 - 13: /SIDS2/gcgdata/geneseq/NA1992.DAT.*
 - 14: /SIDS2/gcgdata/geneseq/NA1993.DAT.*
 - 15: /SIDS2/gcgdata/geneseq/NA1994.DAT.*
 - 16: /SIDS2/gcgdata/geneseq/NA1995.DAT.*
 - 17: /SIDS2/gcgdata/geneseq/NA1996.DAT.*
 - 18: /SIDS2/gcgdata/geneseq/NA1997.DAT.*
 - 19: /SIDS2/gcgdata/geneseq/NA1998.DAT.*
 - 20: /SIDS2/gcgdata/geneseq/NA1999.DAT.*
 - 21: /SIDS2/gcgdata/geneseq/NA2000.DAT.*
 - 22: /SIDS2/gcgdata/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	889.4	99.8	891	22	AAF23529
2	564.4	53.3	917	21	AAF21748
3	314.3	38.5	524	21	AAO01823
4	298.6	33.5	485	21	AAO01823
5	298.6	33.5	485	21	AAO01823
6	273.8	30.7	426	22	AAH2586
7	160.8	18.0	169	16	AAAT1034
8	150.8	16.9	187	21	AAH7147
9	141.2	15.8	865	22	AAH03517
10	141.2	15.8	1229	22	AAH13750
11	141.2	15.8	1364	22	AAH85548

12	108.2	12.1	792	22	AAH45143
13	104	11.7	998	22	AAF59611
14	102.4	11.5	767	22	AAF73810
15	102.4	11.5	858	22	AAH03517
16	102.4	11.5	858	22	AAH03517
17	91.4	10.3	692	21	AAO10889
18	89.8	10.1	662	21	AAO03880
19	89.8	10.1	698	21	AAO03879
20	81	9.1	936	22	AAF58252
21	81	9.1	936	22	AAF58254
22	81	9.1	936	22	AAF58257
23	81	9.1	936	22	AAF58259
24	81	9.1	936	22	AAF58262
25	74.1	8.9	1101	21	AAO01003
26	74.1	8.9	1101	21	AAO01003
27	67.4	7.6	936	22	AAF58252
28	67.4	7.6	936	22	AAF58254
29	67.4	7.6	936	22	AAF58257
30	67.4	7.6	936	22	AAF58259
31	67.4	7.6	936	22	AAF58262
32	67.4	7.6	936	22	AAF58262
33	67.4	7.6	938	22	AAF58255
34	66.4	7.5	128	21	AAO12100
35	64.8	7.3	397	22	AAI14092
36	64.8	7.3	397	22	AAI14092
37	64.8	7.3	397	22	AAI14092
38	58.6	6.6	1370	22	AAH08557
39	58.6	6.6	1250	22	AAI57960
40	58.6	6.6	1551	22	AAI57946
41	56.4	6.3	356	21	AAH30950
42	55.4	6.2	1309	21	AAH89661
43	52.8	5.9	285	22	AAI21436
44	52.8	5.9	285	22	AAI46725
45	52.8	5.9	285	22	AAI07131

ALIGNMENTS

RESULT 1
ID: AAF23529 standard; DNA; 891 BP.
AC: AAF23529;
DT: 22-MAR-2001 (first entry)
DE: Human NADE DNA.
KW: Neurotrophin receptor; p75-NTR; NGF-induced apoptosis;
KW: Neurogenetic disease; NF-kappaB; ds.
XX: Homo sapiens.
XX: MO200075278-A2.
XX: 14-DEC-2000.
XX: 07-JUN-2000; 2000MO-US15621.
XX: 07-JUN-1999; 99US-0327750.
XX: (UYCO) UNIV COLUMBIA NEW YORK.
XX: Sato T.
XX: WPI; 2001-061707/07.
XX: New p75-neurotrophin receptor-associated cell death executor (NADE) and
XX: the gene encoding NADE, useful for modulating the activity of p75NTR
XX: and for detecting neurodegenerative diseases -
XX: Disclosure; Fig 1; 134pp; English.

Matches 565; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 326 ccaggaacacgaagatgagcagccttatgagatgagagagagacccctttggg 385
Db 305 ccaggaacacgaagatgagcagccttatgagatgagagagacccctttggg 364
Qy 386 agagagtgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 445
Db 365 agagagtgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 424
Qy 446 taattttcagatgagcagcagcagcagcagcagcagcagcagcagcagc 505
Db 425 taattttcagatgagcagcagcagcagcagcagcagcagcagcagcagc 484
Qy 506 tgataggaatatcatgagagatgagagagagagagagagagagagagagag 565
Db 485 tgataggaatatcatgagagatgagagagagagagagagagagagagagag 544
Qy 566 gttaggaatatgctgagatgcttcttcttcttcttcttcttcttcttct 625
Db 545 gttaggaatatgctgagatgcttcttcttcttcttcttcttcttcttct 604
Qy 626 tgaattttcagatgagcagcagcagcagcagcagcagcagcagcagcagc 685
Db 605 tgaattttcagatgagcagcagcagcagcagcagcagcagcagcagcagc 664
Qy 686 gctgtttttcttcttcttcttcttcttcttcttcttcttcttcttcttct 745
Db 665 gctgtttttcttcttcttcttcttcttcttcttcttcttcttcttcttct 724
Qy 746 cctatgtattcttcttcttcttcttcttcttcttcttcttcttcttcttct 805
Db 725 cctatgtattcttcttcttcttcttcttcttcttcttcttcttcttcttct 784
Qy 806 ctttgatcagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 865
Db 785 ctttgatcagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 844
Qy 866 aaagcgaataaagcatttcttcttcttcttcttcttcttcttcttcttct 891
Db 845 aaagcgaataaagcatttcttcttcttcttcttcttcttcttcttcttct 870

RESULT 3

AA01005 3
ID AAC01005 standard; cDNA; 532 BP.
XX AC AAC01005;
XX DT 06-OCT-2000 (first entry)
XX DE Human secreted protein 5' EST, SEQ ID NO: 1003.
XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX CS gene therapy; chromosome mapping; as.
XX PN Homo sapiens.
XX EP1033401-A2.
XX PD 06-SEP-2000.
XX PF 21-FEB-2000; 2000EP-0200610.
XX PI 26-FEB-1999; 99US-012487.
XX PA (GEST) GENSET.
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX DR WPI; 2000-500381/45.
XX DR P-PSDB; AAG00999.
XX

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1; SEQ ID 1003; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence; the 5' ESTs were prepared from total human RNAs or poly(A+ RNAs
CC from 3 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region of the mRNA. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
XX vectors.
XX Sequence 532 BP; 151 A; 121 C; 163 G; 91 T; 6 other;

Query Match 38.5%; Score 343; DB 21; Length 532;

Best Local Similarity 98.6%; Pred. No. 6e-93;
Matches 340; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 281 agaaaacacacccagaaaataatctcatctgcaaatattccaccaggaaaacgaaga 340
Db 188 agaaaacacacccagaaaataatctcatctgcaaatattccaccaggaaaacgaaga 247
Qy 341 gatggagcagcctatgagatgagagagagagagagagagagagagagagagagag 400
Db 248 gatggagcagcctatgagatgagagagagagagagagagagagagagagagagag 307
Qy 401 ccagcctcaggaatcagcagcagcagcagcagcagcagcagcagcagcagcagc 460
Db 308 ccagcctcaggaatcagcagcagcagcagcagcagcagcagcagcagcagcagc 367
Qy 461 cataccantagcagatcaatgatggatggatggatggatggatggatggatggat 520
Db 368 cataccantagcagatcaatgatggatggatggatggatggatggatggatggat 427
Qy 521 catggagagatgag 580
Db 428 catggagagatgag 487
Qy 581 gcgtatccttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 625
Db 488 gcgtatccttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 532

RESULT 4

AA023528
XX AA023528 standard; DNA; 700 BP.
XX AC AA023528;
XX DT 22-MAR-2001 (first entry)
XX DE Mouse NADE DNA.
XX KW Neurotrophin receptor; P75-NTR; NGF-induced apoptosis;
XX CS neurogenetic disease; NF-kappaB; ds.
XX XS Mus sp.
XX XN WO200075778-A2.
XX PD 14-DEC-2000.
XX PF 07-JUN-2000; 2000WO-US15621.
XX

Oy 566 gttcaggagattctgctgcttctcattgagggagctctctcattcaccatgaccatcatga 625
 Db 426 kttaggagattctgctgcttctcattgagggagctctctcattcaccatgaccatcatga 485

RESULT 6

AAH87586
 ID AAH82586 standard; cDNA; 426 BP.
 AC AAH82586;
 XX
 DT 25-SEP-2001 (first entry)
 DE Human ovarian tumour associated polynucleotide sequence SEQ ID NO:210.
 KW Human; ovarian tumour; ovarian cancer; diagnosis; gene therapy;
 KW immunogenic; vaccine; ss.
 XX
 XX Homo sapiens.
 PN WO2000151513-A2.
 XX
 PD 19-JUL-2001.
 XX
 PF 16-JAN-2001; 2001WO-US01575.
 XX
 PR 14-JAN-2000; 2000US-0176722.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Algate PA;
 XX
 DR WPI; 2001-425866/45.
 XX
 PT Novel ovarian tumor proteins, and nucleic acids encoding them, used to
 PT treat and diagnose cancers, particularly ovarian cancer -
 PS
 PS Claim 5; Page 100; 338pp; English.
 CC AAH83377 to AAH83878 represent human ovarian tumour-associated
 CC polynucleotide sequences which encode ovarian tumour proteins. The
 CC ovarian tumour protein and polynucleotide sequences have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. The
 CC development of cancer, particularly ovarian cancer, can be used to inhibit
 CC the development of cancer, particularly ovarian cancer. They can also
 CC be used to diagnose the onset and progression of cancer.
 XX
 XX Sequence 426 BP; 86 A; 106 C; 90 G; 137 T; 7 other;

Query Match 30.7%; Score 273.8; DB 22; Length 426;
 Best Local Similarity 97.5%; Pred. No. 3.2e-72;
 Matches 308; Conservative 0; Mismatches 5; Indels 3; Gaps 3;
 Oy 552 cttagggagctgagctgaggaattctcgtcattccttctgagggagctctcattcacc 611
 Db 1 cttagggagctgagctgaggaattctcgtcattccttctgagggagctctcattcacc 60
 Oy 612 catgaccatcatgatgaatttgccttctgcttgcctcctgcattatcatgagatta 671
 Db 61 catgaccatcatgatgaatttgccttctgcttgcctcctgcattatcatgagatta 120
 Oy 672 atactgtattccctgctgtttcttcttcttcttcttcttcttcttcttcttcttctg 731
 Db 121 atactgtattccctgctgtttcttcttcttcttcttcttcttcttcttcttcttctg 179
 Oy 732 cgtttctgtggaacctatgtatttcttcttcttcttcttcttcttcttcttcttcttctg 790
 Db 180 cgtttctgtggaacctatgtatttcttcttcttcttcttcttcttcttcttcttcttctg 238
 Oy 791 ctatttgagagctgcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 850
 Db 239 ctatttgagagctgcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 298

Oy 851 tgcattgaaaaaattta 866
 Db 299 tgcattgaaaaaattta 314

RESULT 7

AAH21034
 ID AAH21034 standard; cDNA to mRNA; 169 BP.
 AC AAH21034;
 XX
 DT 03-JUL-1996 (first entry)
 DE Human gene signature HUMGS02316.
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 XX cell typing; abnormal cell function; ss.
 OS Homo sapiens.
 PN MO9514772-A1.
 XX
 PD 01-JUN-1995.
 XX
 PF 11-NOV-1994; 94WO-JF01916.
 XX
 PR 12-NOV-1993; 93JP-0355504.
 XX
 PA (MATSU/) MATSUDARA K.
 PA (ORUBO/) ORUBO K.
 XX
 PI Matsubara K, Okubo K;
 XX
 DR WPI; 1995-206931/27.
 XX
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues
 XX
 XX Claim 1; Page 786; 2245pp; Japanese.
 CC A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in AAT19001-T26837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constituted so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which the cDNA was
 CC derived. The GS sequences are determined by using a cDNA library made
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.

Sequence 169 BP; 47 A; 30 C; 34 G; 58 T; 0 other;
 Query Match 18.0%; Score 160.8; DB 16; Length 169;
 Best Local Similarity 98.8%; Pred. No. 1.7e-38;
 Matches 162; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 728 gatcccttctgtgacacctatgtatttctcatttctcattgagggctgtgtgtgacag 787
 Db 1 gatcccttctgtgacacctatgtatttctcatttctcattgagggctgtgtgtgacag 60
 Oy 788 ctctctattgagagctgcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 847

Search completed: March 11, 2002, 16:36:43
Job time: 3232 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using av model
Run on: March 11, 2002, 16:33:03 ; Search time 1319.57 seconds
11139-225 Million cell updates/sec

Title: US-09-327-750d-29
Perfect score: 891
Sequence: 1 accaccatccccctctat.....ataaagcaatttaaaagc 891
Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 1472140 seqs, 824889755 residues
Total number of hits satisfying chosen parameters: 2944280
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBank
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.m.*
5: gb.com.*
6: gb.ph.*
7: gb.pl.*
8: gb.pr.*
9: gb.ro.*
10: gb.sts.*
11: gb.un.*
12: gb.vi.*
13: gb.y.*
14: gb.yi.*
15: en.ba.*
16: en.in.*
17: en.in.*
18: en.in.*
19: en.in.*
20: en.in.*
21: en.ov.*
22: en.pat.*
23: en.ph.*
24: en.pl.*
25: en.ro.*
26: en.sts.*
27: en.y.*
28: en.yi.*
29: en.yi.*
30: en.yi.*
31: en.yi.*
32: en.yi.*
33: en.yi.*
34: en.yi.*
35: en.yi.*
36: en.yi.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	899.4	99.8	891	9 HUMOGC	M38188 Human unkno
2	889.4	99.8	891	9 AF187064	AF187064 Homo sapi
3	609.4	68.4	793	9 BC003190	BC003190 Homo sapi
4	363	40.7	854	10 AF097440	AF097440 Mus muscu
5	352.2	32.0	700	10 G13550	G13550 Mus muscu
6	352.2	32.0	700	10 AF187066	AF187066 Mus muscu
7	244.2	29.7	519	10 AF187065	AF187065 Rattus no
8	245.4	27.5	98274	9 HS17487	HS17487 human STS
9	242.4	27.2	244	11 G13550	G13550 human STS
10	185.2	20.8	45678	9 HS17487	HS17487 human STS
11	138.6	17.8	69648	9 HS17487	HS17487 human STS
12	141.2	15.8	1229	9 AK000959	AK000959 Homo sapi
13	141.2	15.8	1364	6 AX100231	AX100231 Sequence
14	141	15.8	43952	9 HS10554	HS10554 Human DNA
15	132.4	14.9	477	11 G35294	G35294 human STS
16	122.2	13.7	35214	9 HS19894	HS19894 Human DNA
17	117.2	13.2	14578	2 AL357046	AL357046 Homo sapi
18	117.2	13.2	14578	2 AL357046	AL357046 Homo sapi
19	117.2	13.2	18297	9 AL353305	AL353305 Homo sapi
20	110.8	12.4	31321	9 HS17487	HS17487 human STS
21	110.8	12.4	41029	2 HS17487	HS17487 human STS
22	109.8	12.3	642	9 AF183416	AF183416 Homo sapi
23	109.8	12.3	828	9 AF220189	AF220189 Homo sapi
24	109.2	12.3	40584	9 AL133348	AL133348 Human DNA
25	108.2	12.1	791	9 AF237783	AF237783 Homo sapi
26	104	11.7	898	6 AX078272	AX078272 Sequence
27	102.4	11.5	744	9 AF251053	AF251053 Homo sapi
28	77	8.6	43952	9 HS10554	HS10554 Human DNA
29	74.6	8.4	1163	9 BC005988	BC005988 Homo sapi
30	74.6	8.4	504	11 G24641	G24641 human STS
31	69.4	7.0	728	10 AF097439	AF097439 Mus muscu
32	59.6	6.7	40822	2 HS17487	HS17487 human STS
33	59.6	6.7	40822	2 HS17487	HS17487 human STS
34	59.4	6.7	138872	2 AC027560	AC027560 Homo sapi
35	59.4	6.7	176039	2 AC073065	AC073065 Homo sapi
36	58.6	6.6	1208	9 AK026349	AK026349 Homo sapi
37	58.6	6.6	91480	9 HSDJ823F3	AL079333 Human DNA
38	58.4	6.6	156275	2 AL513285	AL513285 Homo sapi
39	57	6.4	196367	9 AC020740	AC020740 Homo sapi
40	56.6	6.4	2269	10 AF097437	AF097437 Mus muscu
41	56.6	6.4	188896	2 AC037486	AC037486 Homo sapi
42	56.2	6.3	127606	2 AL356298	AL356298 Homo sapi
43	56.2	6.3	127606	2 AL356298	AL356298 Homo sapi
44	56.2	6.3	127606	2 AL356298	AL356298 Homo sapi
45	56.2	6.3	127606	2 AL356298	AL079303 Human Chr

ALIGNMENTS

RESULT 1
LOCUS HUMOGC
DEFINITION Human unknown protein from clone PHGR74 mRNA; complete cds.
ACCESSION M38188
VERSION M38188.1
KEYWORDS
SOURCE Human ovary
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 891)
AUTHORS Zimmer, M. and Scheit, K.H.
TITLE Characterization of three abundant mRNAs from human ovarian granulosa cells.
JOURNAL J Biol Chem
MEDLINE 91025550
COMMENT Draft entry and computer-readable sequence for [DNA 9, 479-485 (1990)] kindly submitted

Db 1 ACCCCATCCCCACCTCTATACCGGCTCTCCATTTTGGTGGCTGCAACAGCTCTGGGAAG 60
Qy 61 aatcccgaggaacagaaaaatggttgggttggggaagggagggaggaagctggga 120
Db 61 AATCCCGGGAACGAAAAATGCTGGGTGGGGGAAGGAGGTAAGGGAGAAAGCTGGA 120
Qy 121 gggaggggcttaattggagcccgtagagggcgaggaactctaaagtgggaaaa 180
Db 121 GGGAGGGGCTTAATTGGAGGCGCCGTAGAGACGCGGGACCTCTTAAGTGGGAANA 180
Qy 181 acgaaatataaaaccccttggatcatcagggctctgaatcctctgctcagagcaccagc 240
Db 181 ACGAAATATAAAATCTTTGATATCAGGGCTGAATCTCTGCTGTCAGAGCACCAGC 240
Qy 241 attcagctctctcccttggcttcttacttcttctaaagaaaaaacacccagaaaaa 300
Db 241 ATTCAGTCT 300
Qy 301 aaatcctcctcctgagatattatccaggaagaaagagagtgaggcgagcctatgaga 360
Db 301 AAATCTCTCATGCGCAATATTCACCGAGGAAGAGAGATGGAGAGCTTGCAGGA 360
Qy 361 atgagagagagagagcgcccttggggaggggtgaagggcaccagcctcaggaagaatgac 420
Db 361 ATGGAGAGGAGAGCGCCCTTTGGGAGGAGGTGAAGGCCACCGCTGACGGAATCGAC 420
Qy 421 gggagcagggctcggagactgcccctaatcttctgagggccatacccaataagcagatca 480
Db 421 GGGAGCAGGCTCGGCGCTACTGCGCCCTTAATTTTCGATGGGCAATACCCATAGGAGATCA 480
Qy 481 atgtaggagagggag 540
Db 481 ATGATGGAGTGGGTGGAGATGGAGATGATATGGAATATTCATGGAGGAGATGGAGAA 540
Qy 541 tcaagagaaactatggagagcgagtcagtgaggaatctctgctatcctctatggggagc 600
Db 541 TCAGAAGAAACTATAGGAGCTGCGATTCAGGAATCTCTGCGTATCTTTATGGGGAGC 600
Qy 601 tctcctacccctgacatcatgagtaatttgccttactgctgactctgctccttta 660
Db 601 TCTCTATACCATGACATCATGATGATATTTGCTTTATGCTTGACTTCTCCATTTA 660
Qy 661 tcatgagattactatgattcccgctgtttctttcttcttcttcttcttcttcttcttct 720
Db 661 TCATGAGATTAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Qy 721 ctctactgactcgttctgctggaacctatgttattccatgtgtcaagtggctctgtg 780
Db 721 CTTTACTGATCCCTTCTGCTGGAACCTATGTTATTCATGTCATGTCATGTCATGTCAT 780
Qy 781 ttgcagctctctattgagattgcttgccttgcctcagtgtaagtcttctgcagcagtagt 840
Db 781 TTGCCAGCTCTATTGAGATTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Qy 841 ttccaccatttcagcagaaattttaaagcctaagcgaatttaaaagc 891
Db 841 TTCACCCATTTCATGCGAAAAATTTAAAGCTAATAAAGCAATTTAAAAAGC 891

RESULT 3
BC003190 793 bp mRNA PRI 12-JUL-2001
LOCUS Homo sapiens, p75NTR-associated cell death executor; ovarian
cancer cell protein (13kD), clone MGC:802 IMAGE:3357965, mRNA,
complect cds.
ACCESSION BC003190.1 GI:13112030
VERSION
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 793)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (13-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Email: cgas@nci.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
Info@bcsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fyell, Erin Garland, Fan Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Nan Lee, Michael Ling, Carolei Meselson, Andrus Seldin, Steven
P. Ward, Alexander Wang, David Wang, Patrick Ward, Jeff Ward,
Schelin, Duane Smalley, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.
Clone Distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 6 Row: k Column: 22
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 7657043.
Location/Qualifiers
i. 1793am="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:802 IMAGE:3357965"
/tissue_type="eye, retinoblastoma"
/clone_lib="NIH_MGC_16"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
196..531
/codon_start=1
/product="p75NTR-associated cell death executor; ovarian
cancer cell protein (13kD)"
/db_xref="taxon:9606"
/db_xref="GI:13112031"
/translation="MANHOENEMEDPPLGGGGHOPAGNRGGOARRLA
PNFRMAIPNRQINDMGDDDMIFNEMREIRKRLQLRNLRLMGLSNHHD
HHDEFCLMP"
BASE COUNT 225 a 170 c 205 g 193 t
ORIGIN

Query Match 58.4%; Score 609.4; DB 9; Length 793;
E-value Similarity 99.8%; Prev. Name: 1e-15;
Matches 610; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 281 agaaaaaacacacagaaaaaaatctctatcagcgaatattccacaggaacacgaaga 340
Db 165 AGAAAAAACACACAGAAAAAAATCTCATCTGCAAAATATTCCACGAGAAACGAAGA 224
Qy 341 gatggagcgcctatgcagaatgcagaggaagaccgccccttgggagaggtgaagggcca 400
Db 225 GATGGAGCAGCCTATGCAGAAATGCAGAGGAAGACCGCCCTTTGGGAGGAGGTGAAGGCCA 284
Qy 401 ccagcctgcaggaatgcaggggagacaggtcgcgcagcttgccttaatttctgattgggc 460
Db 285 CCAGCTGCAGGAATGCAGGGGACAGGCTGCCCGACTGCCCTTAATTTTCATGGGCGC 344
Qy 461 cataccatagcagcatcatgatggatggatggatggatggatggatggatggatggat 520
Db 345 CATACCAATAGCAGATCAATGATGGGATGGGTGGAGATGGAGATGATGGAATATT 404

Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers to their source databases:
 EMBL: European Molecular Biology Laboratory
 GenBank: National Center for Biotechnology Information
 on the MORGAP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/MORGAP
 CTA-71487 is from the human BAC library described in U-J. Kim et al. (1996) Genomics 34, 213-218.
 VECTOR: pBeloBAC11

This sequence is the entire insert of clone CTA-71487.

FEATURES

source
 1..38274 Homo sapiens
 /organism="Homo sapiens"
 /db_xref="GenBank:U00000"
 /chr_name="22"
 /map="q12.2-13.2"
 /clone="CTA-71487"
 /clone_lib="CT1578SK-A2"
 109..136
 /note="14 copies 2 mer ac 100 conserved"
 183..258
 /note="LMCS repeat: matches 7840..7913 of consensus"
 1177..1446
 /note="match: GSS: Em:AQ373724"
 1301..1356
 /note="match: GSS: Em:AQ373724"
 2568..5038 repeat: matches 2693..2750 of consensus"
 /note="LMOD repeat: matches 3673..6170 of consensus"
 5036..5486
 /note="LMBC repeat: matches 2052..2207 of consensus"
 5496..6885
 /note="LMC repeat: matches 1..1538 of consensus"
 6896..7185
 /note="L2 repeat: matches 2356..2710 of consensus"
 complement(7187..7611)
 /note="match: GSS: Em:AQ566001"
 7369..7394
 /note="MIR repeat: matches 34..256 of consensus"
 complement(7395..8093)
 /note="match: GSS: Em:AQ071449"
 7304..7378
 /note="MIR repeat: matches 20..95 of consensus"
 8086..8300
 /note="MIR repeat: matches 8..255 of consensus"
 8154..8565
 /note="match: GSS: Em:AQ059599"
 8549..8864
 /note="AluX repeat: matches 1..280 of consensus"
 complement(8867..9290)
 9235..9361
 /note="match: GSS: Em:AQ43214"
 /note="L2 repeat: matches 2619..2749 of consensus"
 9378..9890
 /note="match: GSS: Em:AQ284832"
 10135..10428
 /note="AluX repeat: matches 1..294 of consensus"
 10429..10563
 /note="AluX repeat: matches 1..135 of consensus"
 10811..10868
 /note="L2 repeat: matches 2648..2705 of consensus"
 10819..11064
 /note="L2 repeat: matches 2..255 of consensus"
 11112..11321
 /note="MIR repeat: matches 11..255 of consensus"
 11869..12012
 /note="MIR repeat: matches 20..167 of consensus"

12925..13223
 /note="AluY repeat: matches 1..300 of consensus"
 13290..13495
 /note="MER8A repeat: matches 9..214 of consensus"
 13338..13833
 /note="match: GSS: Em:AQ568864"
 13688..13749
 /note="MIR repeat: matches 77..139 of consensus"
 13840..13966
 /note="MIR repeat: matches 22..147 of consensus"
 14162..14245
 /note="MIR repeat: matches 46..136 of consensus"
 14925..15058
 /note="MIR repeat: matches 48..191 of consensus"
 15994..16040
 /note="MIR repeat: matches 102..143 of consensus"
 16116..16184
 /note="MER39 repeat: matches 3..70 of consensus"
 16897..17005
 /note="MIR repeat: matches 9..118 of consensus"
 17618..17700
 /note="MIR repeat: matches 108..192 of consensus"
 17748..17919
 /note="MIR repeat: matches 73..245 of consensus"
 18657..18871
 /note="MIR repeat: matches 35..252 of consensus"
 18927..19230
 /note="AluX repeat: matches 1..301 of consensus"
 19492..19559
 /note="MIR repeat: matches 2..67 of consensus"
 19560..19821
 /note="AluY repeat: matches 34..295 of consensus"
 19822..19959
 /note="MIR repeat: matches 67..212 of consensus"
 19987..20288
 /note="AluY repeat: matches 1..300 of consensus"
 20316..20512
 /note="MIR repeat: matches 82..262 of consensus"
 20513..20741
 /note="LMOD repeat: matches 5970..6224 of consensus"
 20769..21074
 /note="AluSg1 repeat: matches 1..303 of consensus"
 21085..21207
 /note="LM2 repeat: matches 1776..1898 of consensus"
 21212..21391
 /note="L2 repeat: matches 6..178 of consensus"
 21927..22049
 /note="MIR repeat: matches 3..86 of consensus"
 22146..22441
 /note="AluX repeat: matches 1..296 of consensus"
 22526..22610
 /note="L2 repeat: matches 267..349 of consensus"
 22940..23204
 /note="HERV16 repeat: matches 378..650 of consensus"
 23205..23514
 /note="AluX repeat: matches 1..311 of consensus"
 23919..24194
 /note="L2 repeat: matches 36..311 of consensus"
 24456..24590
 /note="L2 repeat: matches 253..349 of consensus"
 24871..24916
 /note="MIR repeat: matches 102..143 of consensus"
 25378..25555
 /note="L2 repeat: matches 2330..2519 of consensus"
 25673..25910
 /note="MIR repeat: matches 4..255 of consensus"
 26722..26798
 /note="L2 repeat: matches 2572..2649 of consensus"
 27040..27107
 /note="L2 repeat: matches 2489..2500 of consensus"
 27108..27157
 /note="AluY repeat: matches 41..304 of consensus"
 27375..27471

```

RESULT 10
LOCUS   HSV351F8      45678 bp      DNA      PRI      23-NOV-1999
DEFINITION   Human DNA sequence from cosmid V351F8, between markers DXS366 and
              DXS87 on chromosome X contains ESTs.
ACCESSION   270719
VERSION     1
KEYWORDS    X
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS     Whiteley M.
TITLE       Direct Submission
JOURNAL     Submitted (09-APR-1995) Sanger Centre, Hinxton, Cambridgeshire,
              CB10 1HQ, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT     WARNING: This sequence is the entire insert of clone V351F8. The
              right end of clone V351F8 is at 45678.
              V351F8 is from the human chromosome X-specific cosmid library.
FEATURES             Location/Qualifiers
     source            1..45678
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="X"
                     /clone="GHC-351F8"
                     /clone_lib="SCCV"
     repeat_region     1..70
                     /note="Alu repeat: matches 80. .1 of consensus"
     repeat_region     642..930
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region     1406..5179
                     /note="L1 element fragment"
     repeat_region     5243..5276
                     /note="17 copies of 2 mer 82 & conserved"
     repeat_region     5279..5567
                     /partial
     repeat_region     9775..10369
                     /note="Alu repeat: matches 304. .1 of consensus"
     repeat_region     10369..10701
                     /note="MSTR element fragment"
     repeat_region     10701..11178
                     /note="MSTR element fragment"
     repeat_region     11178..11978
                     /note="MSTR element fragment"
     repeat_region     11715..11846
                     /note="MSTR element fragment"
     repeat_region     12074..12170
                     /note="MSTR element fragment"
     repeat_region     12227..12486
                     /note="MSTR element fragment"
     repeat_region     12486..12502
                     /note="MSTR element fragment"
     repeat_region     12502..13059
                     /note="MSTR element fragment"
     repeat_region     12756..13059
                     /partial
     repeat_region     14478..14650
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region     14491..14536
                     /note="MSTR element fragment"
     repeat_region     14715..14831
                     /note="MSTR element fragment"
     repeat_region     14899..14933
                     /note="MSTR element fragment"
     repeat_region     15534..15620
                     /note="MSTR element fragment"
     repeat_region     15566..15619
                     /note="31 copies of 3 mer 85 & conserved"
     repeat_region     16352..16436
                     /note="3 copies of 18 mer 98 & conserved"
     repeat_region     16728..16763
                     /note="L1 element fragment"

```

```

repeat_region     16764..17054
                     /partial
     repeat_region     19420..19449
                     /note="Alu repeat: matches 308. .1 of consensus"
     misc_feature     complement(1937..20013)
                     /note="15 copies of 2 mer 87 & conserved"
     repeat_region     20013..20013
                     /note="match: 3' EST N51315 clone 283089"
     repeat_region     20073..20112
                     /note="22 copies of 2 mer 98 & conserved"
     repeat_region     20073..20126
                     /note="10 copies of 4 mer 100 & conserved"
     repeat_region     20073..20126
                     /note="3 copies of 18 mer 87 & conserved"
     repeat_region     23088..23300
                     /note="L1 element fragment"
     repeat_region     23311..23490
                     /note="MTR element fragment"
     repeat_region     23318..23489
                     /note="MTR element fragment"
     repeat_region     23523..23489
                     /note="MTR element fragment"
     repeat_region     23507..24407
                     /note="MTR element fragment"
     repeat_region     24376..24462
                     /note="L1 element fragment"
     repeat_region     24379..24491
                     /note="MSTR element fragment"
     repeat_region     24688..24738
                     /note="MSTR element fragment"
     repeat_region     24810..25016
                     /note="MSTR element fragment"
     repeat_region     25251..25432
                     /note="L1 element fragment"
     repeat_region     25439..25696
                     /note="MER25 element fragment"
     repeat_region     26284..26343
                     /note="MTR element fragment"
     repeat_region     26428..26491
                     /note="MTR element fragment"
     repeat_region     26465..26518
                     /note="MTR element fragment"
     repeat_region     27446..27667
                     /note="L1 element fragment"
     repeat_region     28524..29247
                     /partial
     repeat_region     30160..30325
                     /note="Alu repeat: matches 308. .1 of consensus"
     misc_feature     complement(32825..33230)
                     /note="2 copies of 83 mer 98 & conserved"
     repeat_region     32588..32875
                     /note="match: 5' EST H68599 clone 239077"
     repeat_region     35531..35802
                     /note="match: 3' EST H68239 clone 289077"
     repeat_region     35804..35939
                     /partial
     repeat_region     38625..38705
                     /note="18 copies of 2 mer 83 & conserved"
     repeat_region     38747..38839
                     /note="MTR element fragment"
     repeat_region     38949..39032
                     /note="MTR element fragment"
     repeat_region     38965..39076
                     /note="MTR element fragment"
     repeat_region     41260..41260
                     /note="MTR element fragment"
BASE COUNT  14260 a 9135 c 9327 g 12956 t
ORIGIN

```

Query Match 20.8%; Score 185.2; DB 9; Length 45678;
 Best Local Similarity 64.4%; Pred. No. 5.2e-40;
 Matches 345; Conservative 0; Mismatches 178; Indels 13; Gaps 4;

```
/note="LMD repeat: matches 1357. .1972 of consensus"
11443. .11571
repeat_region
/note="FLAM A repeat: matches 2. .129 of consensus"
11572. .11946
repeat_region
/note="LMD repeat: matches 977. .1357 of consensus"
11947. .12245
repeat_region
/note="AluY repeat: matches 1. .301 of consensus"
12246. .13098
repeat_region
/note="LMD repeat: matches 275. .977 of consensus"
12453. .14534
repeat_region
/note="LMD repeat: matches 17. .113 of consensus"
13523. .13820
repeat_region
/note="MTRD repeat: matches 177. .505 of consensus"
13823. .13932
repeat_region
/note="55 copies 2 mer ct 77% conserved"
13936. .13977
repeat_region
/note="MTRD repeat: matches 201. .241 of consensus"
13978. .14240
repeat_region
/note="AluSg1 repeat: matches 1. .270 of consensus"
14245. .14534
repeat_region
/note="AluSg1 repeat: matches 1. .290 of consensus"
14535. .14921
repeat_region
/note="MTRD repeat: matches 2. .201 of consensus"
14724. .15604
repeat_region
/note="LMD repeat: matches 4646. .5565 of consensus"
15605. .15909
repeat_region
/note="AluSx repeat: matches 24. .312 of consensus"
16069. .16363
repeat_region
/note="AluSx repeat: matches 1. .291 of consensus"
16482. .16641
repeat_region
/note="AluDb repeat: matches 136. .295 of consensus"
16714. .18293
repeat_region
/note="AluSg1 repeat: matches 1. .284 of consensus"
17164. .17280
repeat_region
/note="Lip repeat: matches 4499. .4619 of consensus"
17326. .17377
repeat_region
/note="26 copies 2 mer at 79% conserved"
17704. .17863
repeat_region
/note="AluSc repeat: matches 118. .295 of consensus"
17864. .18175
repeat_region
/note="AluDb repeat: matches 1. .318 of consensus"
18176. .18298
repeat_region
/note="AluSg1 repeat: matches 1. .118 of consensus"
18307. .18379
repeat_region
/note="113 copies 2 mer tg 100% conserved"
18876. .19171
repeat_region
/note="AluSg repeat: matches 1. .295 of consensus"
19183. .19232
repeat_region
/note="25 copies 2 mer ac 100% conserved"
19246. .19533
repeat_region
/note="AluY repeat: matches 5. .292 of consensus"
19539. .19837
repeat_region
/note="MER74B repeat: matches 61. .368 of consensus"
20016. .20274
repeat_region
/note="MER74A repeat: matches 295. .487 of consensus"
20394. .20592
repeat_region
/note="AluSg repeat: matches 1. .302 of consensus"
20701. .20809
repeat_region
/note="LTR1 repeat: matches 682. .785 of consensus"
20877. .21061
repeat_region
/note="LTR1 repeat: matches 1. .195 of consensus"
21186. .21636
repeat_region
/note="LTR29 repeat: matches 3. .498 of consensus"
21609. .21737
repeat_region
/note="MER34 repeat: matches 409. .543 of consensus"
21771. .22076
repeat_region
/note="Lip repeat: matches 2. .308 of consensus"
22878. .22997
repeat_region
/note="MIR repeat: matches 92. .191 of consensus"
23452. .23582
repeat_region
/note="FLAM C repeat: matches 1. .131 of consensus"
23583. .23624
repeat_region
/note="21 copies 2 mer ta 76% conserved"
```

```

23735. .24041
repeat_region
/note="AluSg1 repeat: matches 1. .308 of consensus"
24045. .24072
repeat_region
/note="14 copies 2 mer tt 89% conserved"
25274. .25582
repeat_region
/note="AluDb repeat: matches 1. .289 of consensus"
25583. .25747
misc_feature
/note="match: GSS B63622"
25959. .26012
repeat_region
/note="HERVH21 repeat: matches 23. .79 of consensus"
26144. .26436
repeat_region
/note="AluSg1 repeat: matches 1. .311 of consensus"
27281. .27432
repeat_region
/note="AluDb repeat: matches 137. .289 of consensus"
27436. .27507
repeat_region
/note="LMD repeat: matches 5360. .5434 of consensus"
27536. .27600
repeat_region
/note="LMD repeat: matches 2220. .2283 of consensus"
27601. .27975
repeat_region
/note="MTRD repeat: matches 3. .365 of consensus"
27976. .28273

Query Match      17.8%  Score 158.6;  DB 9;  Length 59648;
Best Local Similarity 60.1%  Pred. No. 1,3e-32;
Matches 409;  Conservative 0;  Mismatches 229;  Indels 43;  Gaps 7;

QY 3 cccatccccccactctatccgcgtctctccattttgtgctgctgcaagctcttgaggaaagaa 62
Db 64012 CCCCTTCCACACGCCCATCTCGCTCTCCATTTTGTCTGCTGCGGAAGCCTGGAGATGGA 64071

QY 63 tcccgaggaaacgaaaaatgctgggtttgggggaggaggagtaaggaggagaaagctggag 122
Db 64072 TC---TACGGGAAATGTGGCTTTCGGGAAGGAGGGCTCGGATTCAGGCGCACCC 64127

QY 123 gagggtcttaattggagccctatcgagagcgcgcggaactctcaaggtgggaaaaa-- 180
Db 64128 CACAGGCGTACTGCTCTCAGGTGGGAAAAAAGTAAGTATTAGATTAAGTCAAGT 64187

QY 181 -----acgaataataaaaatcttggatatacaggggctctgaatcctgctggtcagag 232
Db 64188 CCAGGCGTCTCAATTTAAAGGTGGCTAGTAGGGGCTCTGTCTCAGTGTCTTATCAGTC 64247

QY 233 caccagcattcgtctctctctctctctctctctctctctctctctctctctctctctct 287
Db 64248 CACCAGAAATTCAGCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 64307

QY 288 caaccagaaaaaaatactctctatcgcaaatattcacggagaaacagagagatggag 347
Db 64308 GAATGAGCGCAACATCTCAACGGGGAATAATGCCCAACAGAAACAGAGAGGGGAG 64367

QY 348 cag---cctatgcagaatggagagagacgcgcctttggggagaggtgaagccaccag 404
Db 64368 CAGGCGCTCAGCGCAATGAGAGAAATCCCGCCATTGTGGAGGGGGTGAAAGCCAGAG 64427

QY 405 cctcgaggaaat---cgacggggacaggtcgccgacttgccttaattttcgatgggoc 461
Db 64428 CTGGAGGAATATCAGCGGGGGCGAGTTAGCGGCTGTGCTTATTTTCATTTTCGTGGCC 64487

QY 462 ataccatagcgcccaatgctgggatgggatgggatgggatgggatgggatgggatgg 521
Db 64488 ATACCTAATAGGCATAT-----TGAGCACAAATGAAGCGAGAGATGATGAGAAAGGTTT 64541

QY 522 atggaggagatggaggaalacagaaagaaacttggagagctgagtcagttgaggaaattgct 581
Db 64542 GTAGGGCAGATGATGGAATCAAGAGAAAGACTAGGGAACAGCAGACAGACACTATATG 64601

QY 582 cgtctccttatgggggagctctctaatcaatgacccaatgatgatgaattttgccttat 641
Db 64602 CCTTCTCAAA-----CCTGTACCTGACACCACTATATCTTTTCTCTCTA 64649

QY 642 ccttgactcctgcattatc 662
Db 64650 CTTTGAATCTTAAAGTTTTC 64670
```

```

OY 637 ttatgcttgaactctaccattatc 662
DB 602 TCATACCTGATCTCTAAAGTTTC 627

RESULT 14
LOCUS HSU105G4 43952 bp DNA PRI 23-NOV-1999
DEFINITION DNA sequence from cosmid U105G4, between markers DX5366 and DX587 on chromosome X contains ESTs.
ACCESSION 52846
VERSION 252846.1 GI:1877217
KEYWORDS X.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Burcon,J.
TITLE Direct Sublation
JOURNAL 007-MAR-1997 Sanger Centre, Hinxton, Cambridgeshire,
UK. E-mail enquiries: humquery@sanger.ac.uk
REQUEST: clones@esanger.ac.uk
COMMENT IMPORTANT: This sequence is the entire insert of clone U105G4. The
true left end of clone U105G4 is at 1 in this sequence. The true
right end of clone U105G4 is at 43952.
U105G4 is from the Lawrence Livermore National Laboratory
flow-sorted X chromosome cosmid library LLOXNC01.
LOCATION/Qualifiers
1..43952 Homo sapiens"
/organism="Homo sapiens"
/label="taxon:9606"
/label="name=X"
/label="name=X"
/clone="lib=LLOXNC01"
/clone="LLOXNC01-U105G4"
39..420
/notes="L1MB8 repeat: matches 520..920 of consensus"
repeat_region
1121..1233
/notes="L1MA1 repeat: matches 61..170 of consensus"
repeat_region
1121..1203
/notes="L1MB5 repeat: matches 61..143 of consensus"
repeat_region
3155..3463
/notes="L1MB5 repeat: matches 1..300 of consensus"
repeat_region
/notes="3463"
/notes="26 copies of 4 mer 88 & conserved"
repeat_region
4207..4505
/notes="L1MB2 repeat: matches 301..1 of consensus"
prim_transcript
<4809..>6786
/notes="match: multiple ESTs; match: R68943 W47079 W46779
H2635 H00808; match: H28051 W46978 H13743 W40640 H26464;
match: D61749 H84840 H88611 H24866 W12033; match: N28622
H96650 M40269 H92750 R21391; match: H96654 N30984 R21973
N20969 W78040; match: R67324 R70485 R70575 N27866 W19406;
match: R65743"
7279..7312
/notes="26 copies of 2 mer 84 & conserved"
repeat_region
7315..7632
/notes="L1MB7 repeat: matches 302..2 of consensus"
repeat_region
7661..7738
/notes="L1MB6 repeat: matches 183..261 of consensus"
repeat_region
7881..7933
/notes="L1 repeat: matches 118..174 of consensus"
repeat_region
8874..9296
/notes="L1 repeat: matches 4965..5390 of consensus"
repeat_region
9151..10034
/notes="L1PA2 repeat: matches 1..891 of consensus"
repeat_region
11822..131639
/notes="L1 repeat: matches 1..100 of conserved"
repeat_region
17600..17897
/notes="L1 repeat: matches 299..2 of consensus"
repeat_region
18288..18693
/notes="L1 repeat: matches 3..424 of consensus"
repeat_region

19079..19315
/notes="MIR repeat: matches 262..13 of consensus"
repeat_region
20482..20770
/notes="L1MB1 repeat: matches 292..9 of consensus"
repeat_region
22479..22522
/notes="L1MB2 repeat: matches 48..1 of consensus"
repeat_region
23853..23986
/notes="L1MB3 repeat: matches 2 of 2 mer 97 & conserved"
repeat_region
23756..23855
/notes="L1MB4 repeat: matches 1..131 of consensus;
incomplete repeat"
repeat_region
23866..23958
/notes="L1MB8 repeat: matches 83..1 of consensus"
repeat_region
26423..26480
/notes="29 copies of 2 mer 81 & conserved"
repeat_region
27936..28102
/notes="L1MB1 repeat: matches 372..442 of consensus"
repeat_region
27978..28196
/notes="L1MB5 repeat: matches 572..344 of consensus"
prim_transcript
/notes="match: 5' EST H41148 clone 172618"
29196..29370
/notes="FAM repeat: matches 175..1 of consensus"
repeat_region
29589..29640
/notes="2 copies of 26 mer 98 & conserved"
prim_transcript
30953..33085
/notes="match: multiple ESTs; match: N25344 M52822 H09953
H41228 H52068; match: N21487 H97562 H09131 T10046; match:
H09552 H09187 H4909 N31459"
31131..31135 of TNL000 transposon insertion"
3134..33355
/notes="L1MB1 repeat: matches 304..97 of consensus"
repeat_region
34634..34742
/notes="MIR repeat: matches 101..237 of consensus"
repeat_region
35476..35566
/notes="L1MB2 repeat: matches 1..88 of consensus"
repeat_region
35644..36167
/notes="L1MB2 repeat: matches 75..553 of consensus"
repeat_region
36304..36602
/notes="L1MB3 repeat: matches 299..1 of consensus"
repeat_region
36606..36838
/notes="L1MB5 repeat: matches 870..640 of consensus"
repeat_region
36897..37415
/notes="L1MB5 repeat: matches 490..913 of consensus"
repeat_region
37855..37996
/notes="L1MB2 repeat: matches 285..419 of consensus"
repeat_region
37979..38231
/notes="L1MB3 repeat: matches 428..698 of consensus"
repeat_region
38934..39028
/notes="L1MB6 repeat: matches 472..576 of consensus"
repeat_region
39346..39461
/notes="MIR repeat: matches 235..117 of consensus"
repeat_region
39800..40160
/notes="L1MB2 repeat: matches 414..53 of consensus"
repeat_region
40760..40835
/notes="L1MB5 repeat: matches 1..136 of consensus;
incomplete repeat"
repeat_region
40899..42656
/notes="L1 repeat: matches 3592..5390 of consensus"
repeat_region
42506..43526
/notes="L1MA5 repeat: matches 1..1047 of consensus"
BASE COUNT 13852 a 8286 c 8337 g 13477 t
ORIGIN

Query Match 15.8%; Score 141; DB 9; Length 43952;
Best Local Similarity 24.0%; Pred. NO. 8.e-26;
Matches 494; Conservative 0; Mismatches 345; Indels 76; Gaps 7;
QY 14 ctctatccgctctctcttctgctgcgaagctctggaagagatcccggaac 73
DB 15173 CCCCACTCCAGATCTCTCTCTTTGGTGTCTGCATAGCCCTAGGAGGATC----TACAA 15228

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using av model

Run on: March 11, 2002, 16:09:50 ; Search time 1609.39 seconds
11351937 seqs, 537289261 residues
5949.141 Million cell updates/sec

Title: US-09-327-750d-29

Perfect score: 891
Sequence: 1 accaccatccccctctat.....ataaagcaatttaaaagc 891

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537289261 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

- 1: em_estfun:**
- 2: em_esthum:**
- 3: em_estlin:**
- 4: em_estom:**
- 5: em_estpl:**
- 6: em_estha:**
- 7: em_estro:**
- 8: em_estov:**
- 9: em_hcc:**
- 10: gq_estl:**
- 11: gq_estc:**
- 12: gq_estb:**
- 13: gq_gss:**
- 14: em_gss_fun:**
- 15: em_gss_hum:**
- 16: em_gss_inv:**
- 17: em_gss_pln:**
- 18: em_gss_pro:**
- 19: em_gss_rpd:**
- 20: em_gss_vit:**
- 21: em_gss_other:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
C 1	607.8	68.2	612	10 AW512400	xx74b09.x
C 2	597.0	67.7	615	10 AI193112	AI193112 q669g09.x
C 3	597.4	67.9	616	10 AI278835	AI278835 au65b10.x
C 4	586.4	66.9	626	10 AI278835	AA576958 nm82d04.s
C 5	577.4	64.8	586	11 BF058976	BF058976 7K35g02.x
C 6	577.4	64.8	606	10 AI917016	AI917016 WP72b04.x
C 7	574	64.4	662	10 AI929106	AI929106 n55b04.y
C 8	565.2	63.4	603	10 AA514324	AA514324 XF31g01.x
C 9	564.4	63.3	568	10 AM131584	AM131584 AV709328
C 10	564.4	63.3	574	10 AI075989	AI075989 ov47b05.x
C 11	564.4	63.3	610	10 AV709328	AV709328 hu51f06.x
C 12	562.8	63.2	574	10 BE222704	BE222704

13	562.8	63.2	745	10	AI132971	AI132971
C 14	562.4	63.1	566	11	AV598179	AV598179 en14609.x
C 15	560.8	62.9	566	11	BE732466	BE732466 nmd1811.
C 16	560.8	62.9	720	10	AV701506	AV701506 hu11404.x
C 17	560.4	62.9	566	10	AI1859419	AI1859419
C 18	559.8	62.8	584	10	AI188510	AI188510 qd14f02.x
C 19	559.6	62.8	837	10	BE728687	BE728687 601563980
C 20	557.8	62.6	578	10	AI156951	AI156951 au81e08.x
C 21	557.8	62.6	599	10	AI033984	AI033984 ow13a08.x
C 22	553.2	62.3	506	10	AM088350	AM088350 xd26f03.x
C 23	553.2	62.3	506	10	BE764807	BE764807 60148756
C 24	553.6	62.1	676	11	BE764807	BE764807 60148756
C 25	553.4	62.1	716	11	BE764807	BE764807 60148756
C 26	552.4	62.0	573	11	BF940956	BF940956 hu52f08.y
C 27	551.8	61.9	772	11	BF940956	BF940956 hu52f08.y
C 28	550.8	61.8	858	11	BE733921	BE733921 602418050
C 29	550.4	61.8	690	11	BE733921	BE733921 602418050
C 30	549.8	61.7	589	10	AI188575	AI188575 qd15b05.x
C 31	549.6	61.7	556	10	AA573181	AA573181 nm51h03.s
C 32	549.6	61.7	728	11	BF972848	BF972848 602241127
C 33	548.4	61.7	566	10	AI033728	AI033728 ox89b08.x
C 34	548.4	61.7	566	10	BE408659	BE408659 601304170
C 35	548.8	61.6	879	11	BE408659	BE408659 602435881
C 36	548.5	61.6	701	11	BE403169	BE403169 602773250
C 37	548	61.5	674	11	BE745809	BE745809 602773250
C 38	547.6	61.5	775	11	BE706329	BE706329 602669646
C 39	546	61.3	707	10	BE732777	BE732777 601565256
C 40	545.4	61.2	704	11	BE488926	BE488926 602534944
C 41	543	60.9	570	10	AI887790	AI887790 wml1b02.x
C 42	540.4	60.7	730	11	BE900237	BE900237 601673325
C 43	538.2	60.4	621	10	AI570521	AI570521 tm77f05.x
C 44	537.4	60.3	574	10	AA843808	AA843808 ak09c04.s
C 45	536.6	60.2	562	10	BE727430	BE727430 601560632

ALIGNMENTS

RESULT 1
AW512400/c
LOCUS
DEFINITION
xx74b09.x1 NCI-CGAP-Lym12 Homo sapiens cDNA clone IMAGE:2849369 3' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74 (HUMAN). mRNA sequence.
ACCESSION
AW512400
VERSION
KEYWORDS
EST:7400.1 GI:7150478
SOURCE
human
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 612)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
TITLE
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: rstraus@nci.nih.gov
URL: <http://cgap.nci.nih.gov>
L1: <http://cgap.nci.nih.gov/Genes/ESTs/7400.1/1547-015>
L2: <http://cgap.nci.nih.gov/Genes/ESTs/7400.1/1547-015>
L3: <http://cgap.nci.nih.gov/Genes/ESTs/7400.1/1547-015>
DNA Sequencing by Maxam-Gilbert
Clone Distribution: NCI-CGAP Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml
Image through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml
Seq primer: -40UP from Gibco
High quality sequence, stop: 422.
Location/Qualifiers
1. 612
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="2849369"
/clone_id="NCI-CGAP-Lym12"
/tissue="lymphoma, follicular mixed small and large cell"

```

Oy 821 aegttctcagcagtagtctccaccattgcatggagaaattaaagcaaaagca 880
Db 75 AGTTTCTCTCAGCAGTAGTCTTCCACCATTTGCTGAGGAAATTTCCCTCAATAAGCA 16
Oy 881 attcaaaagc 891
Db 15 ATTTAAAGC 5

RESULT 3
A1928835/c 626 bp mRNA EST 23-AUG-1999
LOCUS a05b10.k1 schneider fetal brain 00004 Homo sapiens cDNA clone
DEFINITION IMAGE:2519611.3 similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
ACCESSION A1928835.1 GI:5664828
VERSION A1928835.1
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S.,
AUTHORS Kriman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin,
W., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
Wu, J., Wyble, B., Wilson, R., and Wilson, R.
TITLE WASHU-NCI Human EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LINC; contact the
IMAGE Consortium (info@image.linc.gov) for further information.
Seq primer: -400P from Gibco
High quality sequence stop: 475.
FEATURES
Location/Qualifiers
1..626
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Schneider fetal brain 00004"
/sex="Male"
/tissue_type="frontal lobe"
/dev_stage="5 months post-conception"
/note="Organ: brain; Vector: pBluescript SK (Stratagene);
Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
adapted from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
5'-GAGAGAGAGAGAGCTCAAGATCTCTTAATTAATTAATCCCTCCCTCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGAGCTCAAGATCTCTTAATTAATTAATCCCTCCCTCC-3'
The library was
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LNCIB-Area
Science Park, Trieste, Italy). " 1 others
BASE COUNT 175 a 144 c 122 g 184 t
ORIGIN

Query Match 67.0%; Score 597.2; DB 10; Length 626;
Best Local Similarity 98.5%; Pred. No. 5.3e-114;
Matches 602; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
Oy 281 agaaaaaacagagaaaaaaattctcatcattggaataattccaccgagaaacgaaga 340

```


Best Local Similarity	97.98, 1.0	Prej. No. 8 2e-110;	Indels	1; Gaps
Matches	594;	Conservative	0;	Mismatches - 12;
QY	285	aaacacccagagaaaaaaatctccatccatccgaatattccacaggaacacagagagagc	344	
Db	606	AAAACACACAGAAAAAAAATCTCATCATGGCAATA-TCACACAGGAAACAGAGAGATG	548	
QY	345	gagcgcctatcgcaatggagagagacgcctctctggagagagtggaagccacag	404	
Db	547	GNACGCCCTATCGAGATATGAGAGAGAACCCCTTTGGAGAGAGTGGAAGGCCACAG	488	
QY	405	ctctcgaggaattcgacgggagcgcgcgcgcgcctgcgcctaattctctgagggcaca	464	
Db	487	CTTCGACGGAATTCGACGGGGACAGCGCTGGCGCGCTTCGCCCTTAATTTTCGATGGCCATA	428	
QY	465	cccaataggccagatcaatgatgggagtggtggagatgggagatgatctgggaattcatcg	524	
Db	427	CCCAATAGGCGAGATCAATGATGAGTGGGTGGAGATGGAGATGATATGGAAATATTCATG	368	
QY	525	gagcgcctatcgcaatggagagagacgcctctctggagagagtggaagccacag	584	
Db	367	GAGGAGATGAGAGAAATCGAAGAGAACTTCGAGGCTTCGATGGAGATATTCGCTTCGCT	308	
QY	585	atctctatggggagcctcccaatcaacatgaccatcatgatgatgaatttcctcatcg	644	
Db	307	ATCCTTTATGGGGAGCTCTCATACCATGACCATCATGATGAATTTGCCTTATGCT	248	
QY	645	tgactcgcctattatcatgagatgaattactgtgattccgcgctctctctctctctctg	704	
Db	247	TGACTCTCGCTCAATTCATGAGATATATGCTGATGATCCGCTGCTTTCTTTCTCTG	188	
QY	705	catttctcctaatactgctctactgactcgtttgctgaggaacatcatgatttcctcatcg	764	
Db	187	CATTTCCTTAATGCTTTTACTGATCGTTTGGTGGTGTGAACCTATGTATTTCCATGCTG	128	
QY	765	ccaagtggtgcttgctgcccagctctctattgaagtatgcttgctgactcagctgaagt	824	
Db	127	TCAGTGGGCTCTGTGTGCCAGCTCTATTGGAAGATGCCCTTGACATCGATGTGTAAGT	68	
QY	825	ctctcgcagctgatttccacatttgctgaggaatttcgaggaatttaagaacataaagaatt	884	
Db	67	TTCTGTGACGATGATTTTCAACCTTTGCTATGGAAATTTTAAAGCAATATTAAGCAATTT	8	
QY	885	aaaaagc 891		
Db	7	AAAAAG 1		
RESULT	7			
LOCUS	A1929106	662 bp	MRNA	EST
DEFINITION	au5810.v1 Schneider fetal brain 00004 Homo sapiens cDNA clone IMAGE:2519611 5', similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74 (HUMAN); mRNA sequence.			22-AUG-1999
ACCESSION	A1929106			
VERSION	A1929106.1	GI:5665070		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	1. Kuhlmann, R. Metaceps; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. (bases 1 to 662).			
AUTHORS	Hillier, L., Allent, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kubacka, T., Lucy, M., Lee, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Scheet, M., Tan, F., Theising, B., White, Y., Wyllie, T., Waterston, R., and Wilson, R.			
TITLE	WashU-NCI human EST Project			
JOURNAL	Unpublished (1997)			
COMMENT	Contact: Wilson RK Washington University School of Medicine 660 South Euclid Avenue, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810			


```

OY 806 ctttgcaactcagtgtaagttcttgcgcagcagtagttccaccatttgcaggaataattt 865
DB 86 ctttgcaactcagtgtaagttcttgcgcagcagtagttccaccatttgcaggaataattt 27

OY 866 aagcgcataaagcaatttttaaaagc 891
DB 26 AAAGCTAATAAGCAATTTTAAAGAGC 1

RESULT 15
BF732486/c 566 bp mRNA EST 30-MAR-2001
LOCUS nael1d11.x1 NCI-CGAP-Ov18 Homo sapiens cDNA clone IMAGE:343948 3'
DEFINITION similar to SW-HG74, HUMAN O00934 OVARIAN GRANULOSA CELL 13.0 KD
PROTEIN HGR74.1, mRNA sequence.
ACCESSION BF732486
VERSION 1
KEYWORDS EST.
SOURCE Human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 566)
REFERENCE 1 (bases 1 to 566)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps@mail.nih.gov
Tissue Procurement: Christopher A. Koskaluk, M.D., Ph.D., Michael
R. Samuels, Ph.D., Ph.D. cDNA library preparation: M. Benito
S. Samuels, Ph.D., Ph.D. cDNA library construction: The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 453.
FEATURES
source
1..566
/organism="Homo sapiens"
/db_xref="GeneID:343948"
/clone_lib="NCI-CGAP-Ov18"
/tissue_type="fibrotheca"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: Ovary; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site:1: Not 1; Site:2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCACTCTGAAGTGGAGCGCCGCCGACATTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library
was constructed from normal ovarian granulosa cells
constructed by Benito Samuels and M. Fatima Bonaldo."
BASE COUNT 162 a 140 c 110 g 153 t
ORIGIN

```

Search completed: March 11, 2002, 16:09:59
Job time: 10894 sec

```

Query Match 62.9% Score 560.8; DB 11; Length 566;
Best Local Similarity 99.5%; Pred. No. 1.9e-106;
Matches 562; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 327 caggaaacagagatgagcagcttatgcgaatggagagagaccgccctttggga 386
DB 566 CAGGAACCGAGAGATGGAGAGCCTATCGAGATGGAGAGAGAGCCCTTTGGGA 507

OY 387 ggaagtgagggccagcctcgaggaatcgagggagagcagctcgccactgccct 446
DB 506 GAAGGTGAAGGCCACCAGCCTCGACGAATCGAGCGGAGAGCTCGCCGACTTGCCT 447

```


TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 861-8540
 TELEFAX: (617) 861-8540
 INFORMATION FOR SEQ ID NO: 102:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 40328 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-742-185-102

Query Match 5.0%; Score 44.4; DB 3; Length 40328;
 Best Local Similarity 51.0%; Pred. No. 0.012;
 Matches 105; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
 Oy 684 cggctgtttctttcttctgatttcttaataatgcttttaagatcgttctgctgga 743
 Db 20178 CCCCTGCCCTTTTTTTTTCCTTCCTGATGCTTCCTGCATCTTCTTTATTGGA 20237
 Oy 744 accctatgtattcttcctgaagggtgtgtgctgcagcttcttattgaagt 803
 Db 20238 GCTATGTGTGCTCTCTCAGCGTGAATGGTTCCTGATACGACACTGATGGTCTT 20297
 Oy 804 cgttgactcagtgtaagtttctctcagcagtagtttcaaccatttgatgtaaaat 863
 Db 20298 GACTCTTATCCAGTTGGCAGCTCTGTCTTTTAAATGGACATTTAGTCCATTCAC 20357
 Oy 864 ttaagccataaagcaatttaaaaa 889
 Db 20358 TTAAGTTAATATCTTATGATGAA 20383

RESULT 4
 US-08-416-478A-1/c
 Sequence 1, Application US/08416478A
 Patent No. 5755757
 GENERAL INFORMATION:
 APPLICANT: Hercend, Thierry
 TITLE OF INVENTION: New Proteins Produced by Human
 TITLE OF INVENTION: Lymphocytes, DNA Sequences Encoding These Proteins And
 TITLE OF INVENTION: Their Pharmaceutical And Biological Uses
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BROMDY AND NEIMARK, P.L.L.C.
 STREET: 419 Seventh Street, N.W., Suite 400
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20004
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/416,478A
 FILING DATE: 04-APR-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/854,644
 FILING DATE: 08-SEP-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: YUN, Allen C.
 REGISTRATION NUMBER: 37,971
 REFERENCE/DOCKET NUMBER: HERCEND-1A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 628-5197
 TELEFAX: (202) 737-3528
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1871 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 231..1724

LENGTH: 1871 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 231..1724
 US-08-416-478A-1
 Query Match 4.2%; Score 37.2; DB 1; Length 1871;
 Best Local Similarity 57.9%; Pred. No. 0.29;
 Matches 66; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
 Oy 14 ctcctatccggtctctccattttgtgctgcgaagctcttggaaagaatcccggaac 73
 Db 239 CCCCACATCTCTCTATGCTGTGGCAGCCCTCCCAAAAGAGGTGCAAAAG 180
 Oy 74 gaaatgtgcttggaggaagagagagagagagagagagagagagagagagagag 127
 Db 179 GAAAGGGGCGAGTGGTGGGGGTAAAGAGAGACTTCTGCGAGAGAGG 126
 RESULT 5
 US-08-474-988B-1/c
 Sequence 1, Application US/08474988B
 Patent No. 5874250
 GENERAL INFORMATION:
 APPLICANT: Hercend, Thierry
 TITLE OF INVENTION: New Proteins Produced by Human
 TITLE OF INVENTION: Lymphocytes, DNA Sequences Encoding These Proteins And
 TITLE OF INVENTION: Their Pharmaceutical And Biological Uses
 NUMBER OF SEQUENCES: 11
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BROMDY AND NEIMARK, P.L.L.C.
 STREET: 419 Seventh Street, N.W., Suite 400
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20004
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/474,988B
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 09/416,478
 FILING DATE: 04-APR-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/854,644
 FILING DATE: 08-SEP-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: YUN, Allen C.
 REGISTRATION NUMBER: 37,971
 REFERENCE/DOCKET NUMBER: HERCEND-1B
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 628-5197
 TELEFAX: (202) 737-3528
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1871 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 231..1724


```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,967
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-676-967-2

Query Match      3.8%; Score 34; DB 1; Length 2277;
Best Local Similarity 29.0%; Pred. NO. 2.8;
Matches 54; Conservative 38; Mismatches 94; Indels 0; Gaps 0;

QY 572 gaattgtctgctatccttcttggggagctcttcaataccatgaccatcatgatgaatt 631
DB 2224 RNARYTTTGTTTATGTCACNARNYTGRTTAAACKNGTYTCNGTYTFTTNCCT 2165
QY 632 ttcccttctgctgactgcctgccttcatcatgagatttaatactgctgctgcgttt 691
DB 2164 TNGCYTTTTCNCKSNACATGTCNSNSNARNYTGTYGTYTTCYGTTCCTCAVGTGT 2105
QY 692 tctctttctctgattctctcaataatgccttcttctgctgctgctgctgacccctatg 751
DB 2104 TDATYTGNGGYTTNGGYTTTNGGRTGNACNGGYTTNACYTTCCTTTCNCKNARNC 2045
QY 752 tttatt 757
DB 2044 KDATYTT 2039

RESULT 14
US-08-676-974-2/c
; Sequence 2, Application US/08676974
; Patent No. 5770422
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,974
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-09-098-487-2
```

```

; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-676-974-2

Query Match      3.8%; Score 34; DB 1; Length 2277;
Best Local Similarity 29.0%; Pred. NO. 2.8;
Matches 54; Conservative 38; Mismatches 94; Indels 0; Gaps 0;

QY 572 gaattgtctgctatccttcttggggagctcttcaataccatgaccatcatgatgaatt 631
DB 2224 RNARYTTTGTTTATGTCACNARNYTGRTTAAACKNGTYTCNGTYTFTTNCCT 2165
QY 632 ttcccttctgctgactgcctgccttcatcatgagatttaatactgctgctgcgttt 691
DB 2164 TNGCYTTTTCNCKSNACATGTCNSNSNARNYTGTYGTYTTCYGTTCCTCAVGTGT 2105
QY 692 tctctttctctgattctctcaataatgccttcttctgctgctgctgctgacccctatg 751
DB 2104 TDATYTGNGGYTTNGGYTTTNGGRTGNACNGGYTTNACYTTCCTTTCNCKNARNC 2045
QY 752 tttatt 757
DB 2044 KDATYTT 2039

RESULT 15
US-09-098-487-2/c
; Sequence 2, Application US/09098487
; Patent No. 5917025
; GENERAL INFORMATION:
; APPLICANT: COLLINS, Kathleen
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,487
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-09-098-487-2
```


33 rgluproalaValalaLeuileSerGlualeGlyLysAsnCysAlaPro 49
264 GGGAGCC...TTGGCCCTCCCTTGGATGCTGGTGAATAGTGTGCT 310
50 ArgGlyValArgArgPheArgValArgGlnProIleAlaHisTyr 66
311 AGAGGAAATCGTAGCGGTTTCGGCTTAGCGACGCCATCTCGCAGTATAG 360
66 gtrpAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGlu 83
361 ATGGGATATGATGATAGCTTGGAGTTCGAGACCCACAGCAGGATGAGAGAG 410
83 luAnValGlnArgPheGlyArgValArgGlnLeuMetGluLysLeu 99
411 ACAATATGAAAGATGGGGAGGGGTGAGACAGCTGATGGAAGCTG 460
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 116
461 AGGAAAGACAGTGTAGTCTAGTCTGGGGCAGTACAGCAGTACGACCC 510
116 ohHisAspHisAspGluPheCysLeuMetPro 128
511 TCNCCATGCCCTCATGATGATGATGATGATGATGATGATGATGATGAT 547
seq_name: /SDS2/gcdata/geneseq/geneseq/NA2001.DAT:AAH75810
seq_documentation_block:
ID_AA75810 standard; cDNA; 767 BP.
XX AAH75810;
XX 17-OCT-2001 (first entry)
XX Human X chromosome linked gene expression protein 14 coding sequence.
XX Human; X chromosome linked gene expression protein 14; cancer;
XX HIV infection; cytostatic; anti-HIV; chromosome X; ss.
XX Homo sapiens.
XX CN1296969-A.
XX 30-MAY-2001.
XX 23-NOV-1999; 99CN-0124078.
XX 23-NOV-1999; 99CN-0124078.
XX (SHAN-) SHANGHAI BORONG GENE DEV CO LTD.
XX Mao Y, Xie Y;
XX WPI: 2001-48397/53.
XX P-PSDB: AAG66407.
XX Polypeptide-human X chromosome linked gene expression protein 14 and
XX polynucleotide for coding said polypeptide -
XX Claim 6; Page 24 (Disclosure); 31pp; Chinese.
XX The present sequence is the coding sequence for human X chromosome linked
XX gene expression protein 14. The protein and coding sequence are useful
XX for treating diseases e.g. cancer and HIV infection.
XX Sequence 767 BP; 224 A; 143 C; 215 G; 185 T; 0 other;
alignment_scores:
Quality: 450.00 Length: 129
Ratio: 3.879 Gaps: 2
Percent Similarity: 89.922 Percent Identity: 66.667

alignment_block:
US-09-327-750D-30 x AAH75810
Align seg 1/1 to: AAH75810 from: 1 to: 767
1 MetGluSerLysAspGln...GlyValLysAsnLeuAsnMetGluAsn 16
101 ATGGAGTCCAGAGAGAACGAGCTTAAACACATCTCATCTGCGGAAATGT 150
16 PHisGlnLysValGluGluGluGluGluGluGluGluGluGluGluGlu 33
151 CAACGAGGAAATGATGAAAGATGAAAGAGCAAGTTGCTAATAAG 200
33 rgluproalaValalaLeuileSerGlualeGlyLysAsnCysAlaPro 49
201 GGGAGCC...TTGGCCCTCCCTTGGATGCTGGTGAATAGTGTGCT 247
50 ArgGlyValArgArgPheArgValArgGlnProIleAlaHisTyr 66
248 AGAGGAAACCTTAGCGGTTTCGGCTTAGCGACGCCATCTCGCAGTATAG 297
66 gtrpAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGlu 83
298 ATGGGACATATGATGATGATGATGATGATGATGATGATGATGATGAT 347
83 luAnValGlnArgPheGlyArgValArgGlnLeuMetGluLysLeu 99
348 AGAATATGGAAGATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 397
398 AGGAAAGACAGTGTAGTCTAGTCTGGGGCAGTACAGCAGTACGAC 447
116 ohHisAspHisAspGluPheCysLeuMetPro 128
448 TCACCATGACCATCAGCAGTGTGCTTGCCTTATGCCCC 484
seq_name: /SDS2/gcdata/geneseq/geneseq/NA2001.DAT:AAI58581
seq_documentation_block:
ID_AA158581 standard; cDNA; 862 BP.
XX AA158581;
XX 22-OCT-2001 (first entry)
XX Human polynucleotide seq ID NO 784.
XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX Homo sapiens.
XX WO200153312-A1.
XX 26-JUL-2001.
XX 26-DEC-2000; 2000MO-US34263.
XX 21-JAN-2000; 2000US-0488725.
XX 23-APR-2000; 2000US-0524317.
XX 09-JUL-2000; 2000US-0520312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0652191.
XX 19-OCT-2000; 2000US-0693036.
XX 26-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
PA

/gene="Beal"
 /codon_start=1
 /product="Beal protein"
 /protein_id="AA047168.1"
 /db_xref="GI:5702156"
 /translation="MESKDGKYNLNNDHOKKEEKPODITIRREPAVLTSEAG
 KICAPGRRRRFROPFASHYRMDLQVGPQGRREENVRFQGVQVRLMEKLRR
 QLSHSLAVSYDPPPHDHDFCLAP"
 BASE COUNT 563 a 538 c 705 g 447 t
 ORIGIN

alignment_scores: Quality: 683.00 Length: 128
 Ratio: 5.378 Gaps: 0
 Percent Similarity: 99.219

alignment_block:

US-09-327-750D-30 x AF097437

Align seg 1/1 to: AF097437 from: 1 to: 2269

1 MetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspH1 17
 1268 ATGGAGTCCAAAGATCAAGCGCTGGGAAATCTCAACATGGAGATGACCA 1317
 17 sGlnLysLysGluLysGluLysGluLysPFGGlnAspThrIleArgG 34
 1318 TCAGAAAG 1367
 34 LuProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50
 1368 AGCCAGCTGTGGCTGTGACCTCGAGGCTGGCAAACTGTGCCTTGA 1417
 51 GlyIleArgArgPheArgValArgGlnProIleAlaHisTyrArgTT 67
 1418 GGAGTGCAGAGCGGTTCGGGTTCGCGACCCATCGCTCACTATAGATG 1467
 67 pAspLeuMetGlnArgValGlyLysGluLysGluLysGluLysGluLys 84
 1468 GACCTGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1517
 84 sNValGlnArgPheGlyValArgGlnLeuMetGluLysLeuArg 100
 1518 ACCTACAGAGTTTGGGGGTGATGTGAGACACTCATGGAGAGCTGAGG 1567
 101 GluArgGlnLeuSerHisLeuArgAlaValSerThrAspProProH1 117
 1568 GAAGGCGAGCTGAGCCACAGCTGGGGGGGTAGCACTGACCGCCCTCA 1617
 117 sHisAspHisHisAspGluPheCysLeuMetPro 128
 1618 TCATGACCATCATGATGATGATGATGATGATGATGATGATGATGATG 1651

seq_name: gb_ro:BC003254

seq_documentation_block:
 LOCUS BC003254 2901 bp mRNA ROD 20-FEB-2001
 DEFINITION Mus musculus, similar to dentatorubral pallidoluysian atrophy,
 clone MGC:5758, mRNA, complete cds.
 ACCESSION BC003254
 VERSION BC003254.1 GI:13096906
 KEYWORDS MGC.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 REFERENCES
 1 Murayama, Y., et al. 1996. Dentatorubral pallidoluysian atrophy: A
 new form of Batten's disease. J. Inher. Metab. Dis. 19:103-110.
 2 Strausberg, P. et al. 2001.
 TITLE Direct Submission
 SUBMITTER National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

REMARK

NIH-MGC Project URL: http://mgc.nci.nih.gov
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Jeffrey Green M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 cDNA Sequencing by: Baylor College of Medicine Human Genome
 Research Center
 Center code: BCM-HGSC
 Web site: http://www.hgsc.bcm.tmc.edu/cdna/
 Contact: villalobos@bcm.tmc.edu.
 Villalobos, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
 A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
 Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Series: IRAP Plate: 9 Row: 0 Column: 6.

FEATURES

Source

Location/Qualifiers
 2901
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="MGC:5758 IMAGE:3500522"
 /tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating
 ductal carcinoma, 5 month old virgin mouse."
 /clone_lib="NCI-CCAP_Mam6"
 /lab_host="DH108"
 /note="Vector: PCMV-SPOrt6"
 755..2359
 /codon_start=1
 /product="similar to dentatorubral pallidoluysian atrophy"
 /protein_id="AA047168.1"
 /db_xref="GI:13096907"
 /translation="MVQFLLPLPLPLKPPHVHTLHPRAKHEHTPPSHOSLOSP
 PQLPPLSPPAATGTCPLPLTATQIKQPAEYEPPEPVPVPAARSPPPKVVDVPS
 HASQSAFNKHLQFGNSCARLDYFVPLEKSLAKRADLVKVRREAEQARAEKE
 RRREREREKEREREKELERESVLAQGRAPVECVSLGVPHPRPPEGSAVATVPP
 YLGPDTALRTLSEYARHYVSPQRNHPFVPLGAVDPGLGLGTYNVALYSSDPAARE
 REREREROLRLKAPGEPKPSLEPLHPVPGGLDFFPHRGGLALQPGPGHFFP
 PPSLGPRLERLALAGPALPDMSTYARLAAERQHAERVAALGNDPLARLQMLNVT
 PHHQSHLHSHLHQDHAISASVYRDIQPLASGSHLRIPTTPACTLPNPLLPH
 LHLQRLAAKAPRDELPASLSTKATPMSAAHQLOQMHASLQRLALEQQQWLHAH
 HFLSTPFADEGSHKRESKST

CDS

623 a 958 c 791 g 529 t
 BASE COUNT
 ORIGIN

alignment_scores:

Quality: 619.00 Length: 118
 Ratio: 5.291 Gaps: 0
 Percent Similarity: 99.153

alignment_block:

US-09-327-750D-30 x BC003254

Align seg 1/1 to: BC003254 from: 1 to: 2901

1 MetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspH1 17
 223 ATGGAGTCCAAAGATCAAGCGCTGGGAAATCTCAACATGGAGATGACCA 272
 17 sGlnLysLysGluLysGluLysGluLysPFGGlnAspThrIleArgG 34
 273 TCAGAAAGG 322
 34 LuProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50
 323 AGCCAGCTGTGGCTGTGACCTCGAGGCTGGCAAACTGTGCCTTGA 372
 51 GlyIleArgArgPheArgValArgGlnProIleAlaHisTyrArgTT 67
 373 GGAGTGCAGAGCGGTTCGGGTTCGCGACCCATCGCTCACTATAGATG 422

Sequence 858 BP: 236 A; 175 C; 241 G; 205 T; 0 other;
 Sequence 858 BP: 236 A; 175 C; 241 G; 205 T; 0 other;

alignment_block:

total human RNAs or poly(A)⁺ RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs, and in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNA sequences. They are used in diagnostic, forensic, gene therapy, and genome mapping procedures. They are used to identify regulatory sequences and to design expression vectors.

```

alignment_scores:      Length: 125
      Quality: 429.50
      Ratio: 3.940
      Gaps: 1
      Percent Similarity: 87.200
      Percent Identity: 64.000

alignment_block:
US-09-327-750D-30 x AAC10889      ..

Align seg 1/1 to: AAC10889 from: 1 to: 692

4  LysaspGlnGlyValLysAsnLeuAsnMetGluAsnAspHisGlnLysLys 20
      AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
249  AGAGAGAGAGGGCTTAACACATCTCATCGTGAAGAAATGTCACAGGAGAA 298
      AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
29  sGlnGlyLysGlnGlyLysProGlnAspThrIleArgAGLProAlay 37
      AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
297  TGTGTAAGAAAGTGAAGAGGACAGTGTCTTAATAAAGGGGAGCC...T 345
      AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
3  alalaLeuLysSerGluAlaGlyLysAspCysAlaProArgGlyGlyArg 53
      AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
346  TGGCTCTACCTCTTGTCTGTGAATCTATGCTGTGCTCTAGAGAAATCTGT 395
      AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

```

seq name: /SIDS2/acadata/geneseq/geneseq/NA2000.DAT:AAC10889

54 A r g P h e A r g V a l A r g I n P r o I l e A l a H i s T y r A r g T r p S p L e u M e 70
396 A G C G T T C C G G T T A G C A C C C A T C T C A T A T T A T A G A T G G A T A T 445
70 T G L A r g V a l G L y C u p T r p G l n G L y A r g M e t A r g G L y A s n V a l G l n A 87
446 G C A T A G C T T G G A G A C C A C C A G C A G C A G A T G A N A A G A G A A T A T G G A A 495
87 r p h e G l y G l y A s p V a l A r g I n L e u M e t G l y L y L e u A r g G l u A r g G l n 103
496 G G A T T G G G A G A G A G T T A G A C A C T G A T G A A A G C T P r o G l i s H i s A s n G 120
104 L e u G l i n S e r L e u A r g V a l L e u T h r A s p P r o P r o G l i s H i s A s p H 145
546 T T C A T A T T C T T G G G A T C A G C A C T A C T A C T A C A C C C C C C A C C A T A G C A 595

104 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspHis 120
|||||
545 ~~TTG~~GAGTCATAGCTGTGGGGCAGTCAGCACTGACCCCTCCACCATGACCA 595

```
1 MetGluSerLysAsp...GlnGlyValLysAsnLeuAnMetGluAsnAs 16
24358 ATGGATCCCAAGAGAAACAGCAGTAAACAGCTCTCAGCATGGAAATGCC 24309
16 pHisGlnLysLysGluGluLysGluGluLysProGlnAspThrIleArgA 33
24308 CACCCAGAAATATGAGAAAGAG...CAAGTTGCTTAATAAAG 24268
33 TGGTCTCAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 49
24267 GGGAGGCC...TTGGCCCTCCCTTTGGATGCTGTGTAATCTGTGCT 24221
50 ArgGlyValArgGlyArgGlyArgGlyValArgGlnProIleAlaHisTyr 66
24220 AGAGAAATCTAGGCGGTTCCTGCTAGGCGAGCCATCTGCGATATAG 24171
66 gTTPAspLeuMetGlnArgValGlnGluGluLysProGlnArgMetArgGlu 83
24170 ATGGATATGCTAGGATGCTAGGATGCTAGGATGCTAGGATGCTAGGATG 24121
83 LuAsnValGlnArgPheGlyValArgGlnLeuMetGluLysLeu 99
24120 AGAATATGGAAGGATGCGGAGGAGGTGAGACAGCTGATGGAAGAAGCTG 24071
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
24070 AGGAAAGCACTGAGTCTAGTCTGCGGCGAGCAGCAGCAGCAGCAGC 24021
116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128
24020 TCACCTGACCATCATGATGAGTTTGCCTTATGCC 23984
seq_name: gb_pr:AF237783
seq documentation block: 791 bp mRNA PRI 02-SEP-2000
LOCUS AF237783
DEFINITION Homo sapiens brain-expressed protein BEX1 (BEX1) mRNA, complete cds.
ACCESSION AF237783
VERSION AF237783.1 GI:9963898
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 791)
AUTHORS Yang, Q.S., Ying, K., Xie, Y., and Mao, Y.M.
TITLE A Novel Human X-linked Brain Expressed Protein
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 791)
AUTHORS Mao, Y.M., Xie, Y., Yang, Q.S., Wu, H., Lin, S. and Ying, K.
TITLE Direct Submision
JOURNAL Submitted (21-FEB-2000) Genetic Research Institute, Fudan University, 220 Handan Rd., Shanghai 200433, P.R.China
FEATURES
source
location/qualifiers
1..791
/db_xref="taxon:9606"
/chromosome="X"
/map="between DXS990 and DX1059 by RH (Stanford G3)*
1..791
/gene="BEX1"
172..549
/gene="BEX1"
/codon_start=1
/product="brain-expressed protein BEX1"
/protein_id="AAC38982.1"
/db_xref="GI:9963898"
/VENUS=RRFVPTPLVYRHWLHGLRQARNRNENRIGEGVRLMKLRLKQLS
HSLRVSSTDPHHDHDFCLMP*
BASE COUNT 214 a 172 c 218 g 187 t
ORIGIN
```

```
alignment_scores:
Quality: 450.50 Length: 129
Ratio: 3.952 Gaps: 3
Percent Similarity: 88.372 Percent Identity: 68.992
alignment_block:
US-09-327-750D-30 x AF237783
Align seg 1/1 to: AF237783 from: 1 to: 791
1 MetGluSerLysAspGln...GlyValLysAsnLeuAnMetGluAsnAs 16
172 ATGGATCCCAAGAGAAACAGCAGTAAACAGCTCTCAGCATGGAAATGCC 221
16 pHisGlnLysLysGluGluLysGluGluLysProGlnAspThrIleArgA 33
222 CACCCAGAAATATGAGAAAGAG...CAAGTTGCTTAATAAAG 262
33 TGGTCTCAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 49
263 GGGAGGCC...TTGGCCCTCCCTTTGGATGCTGTGTAATCTGTGCT 309
50 ArgGlyValArgGlyArgGlyArgGlyValArgGlnProIleAlaHisTyr 66
310 AGAGAAATCTAGGCGGTTCCTGCTAGGCGAGCCATCTGCGATATAG 359
66 gTTPAspLeuMetGlnArgValGlnGluGluLysProGlnArgMetArgGlu 83
360 ATGGATATGCTAGGATGCTAGGATGCTAGGATGCTAGGATGCTAGGATG 409
83 LuAsnValGlnArgPheGlyValArgGlnLeuMetGluLysLeu 99
410 AGAATATGGAAGGATGCGGAGGAGGTGAGACAGCTGATGGAAGAAGCTG 459
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
460 AGGAAAGCACTGAGTCTAGTCTGCGGCGAGCAGCAGCAGCAGCAGC 509
116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128
510 TCACCTGACCATCATGATGAGTTTGCCTTATGCC 546
seq_name: gb_pr:AF251053
seq documentation block: 744 bp mRNA PRI 15-APR-2001
LOCUS AF251053
DEFINITION Homo sapiens X-linked protein mRNA, complete cds.
ACCESSION AF251053
VERSION AF251053.1 GI:13625167
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 744)
AUTHORS Mao, Y., Xie, Y., Zhou, Z., Zhao, W., Zhao, S., Wang, W., Huang, Y., Wang, S., Tang, R., Chen, X. and Wu, C.
TITLE Direct Submission
JOURNAL Submitted (29-MAR-2000) Institute of Genetics, School of Life Sciences, Fudan University, 220 Handan Road, Shanghai 200433, P.R. China
FEATURES
source
location/qualifiers
1..744
/organism="Homo sapiens"
/db_xref="taxon:9606"
101..487
/codon_start=1
/protein_id="AAK34943.1"
/protein_id="AAK34943.1"
/db_xref="GI:13625168"
/translation="MESKERALNNLVNNOENDEKDEKQVANKGEPLALPLNYS
```


/chromosome="X"
 /map="X"
 /clone="GHC-870H8"
 /clone_lib="SCCV"
 475..701
 /note="LI element fragment"
 repeat_region 802..988 element fragment"
 repeat_region 921..1136 element fragment"
 misc_feature 2371..2378
 /note="match: multiple ESTs"
 repeat_region 4189..4246
 /note="29 copies of 2 mer 91 & conserved"
 repeat_region 7496..7874
 /note="LI element fragment"
 repeat_region 8022..8108
 /note="LI element fragment"
 repeat_region 8302..8493
 /note="LI element fragment"
 repeat_region 8585..8671
 /note="LI element fragment"
 repeat_region 8782..9016
 /note="LI element fragment"
 repeat_region 9259..9384
 /note="LI element fragment"
 repeat_region 9624..9731
 /note="MLTIC element fragment"
 repeat_region 9781..9948
 /note="MLTIC element fragment"
 repeat_region 9890..9948
 /note="MLTIC element fragment"
 repeat_region 10181..10392
 /note="MLTIC element fragment"
 repeat_region 10593..10688
 /note="Alu repeat: matches 298..108 of consensus"
 /partial
 /note="Alu repeat: matches 96..1 of consensus"
 repeat_region 11112..11403
 /note="Alu repeat: matches 1..308 of consensus"
 repeat_region 11922..12004
 /note="LI element fragment"
 repeat_region 12021..12208
 /note="LI element fragment"
 repeat_region 12423..12707
 /partial
 /note="Alu repeat: matches 302..1 of consensus"
 repeat_region 12710..12916
 /note="LI element fragment"
 repeat_region 12979..13081
 /note="LI element fragment"
 repeat_region 13473..13532
 /note="MLT21 element fragment"
 repeat_region 15871..15928
 /note="LI element fragment"
 repeat_region 16260..16284
 /note="LI element fragment"
 repeat_region 18778..19056
 /note="Alu repeat: matches 1..308 of consensus"
 repeat_region 22279..22250
 /note="LI element fragment"
 repeat_region 22449..22895
 /note="LI element fragment"
 repeat_region 23239..23374
 /note="LI element fragment"
 repeat_region 23371..23606
 /note="MLT18 element fragment"
 repeat_region 23759..23784
 /note="MLT18 element fragment"
 repeat_region 23705..23778
 /note="MLT18 element fragment"
 repeat_region 23805..23960
 /note="MLT18 element fragment"
 repeat_region 23805..23960
 /note="LI element fragment"

repeat_region 24011..24541
 /note="LI element fragment"
 repeat_region 24638..24770
 /partial
 /note="Alu repeat: matches 150..1 of consensus"
 repeat_region 24829..30975
 /note="LI element fragment"
 BASE COUNT 9462 a 6147 c 6129 g 5583 t
 ORIGIN
 alignment_scores:
 Quality: 450.00 Length: 129
 Ratio: 3.879 Gaps: 2
 Percent Similarity: 89.922 Percent Identity: 66.667
 alignment_block:
 US-09-327-750D-30 x HSV870H8 ..
 Align seg 1/1 to: HSV870H8 from: 1 to: 31321
 1 MetGluSerLysAspGln...GlyValLysAsnLeuAsnMetGluAsnAs 16
 |||||.....
 3508 ATGGAGTCCAAAGAGGACGCGCTTAAACAATCTCATCTGCGGAATGT 3557
 |||||.....
 16 pHISGlnLysLysGluGluLysLysProGlnAspThrIleArgA 33
 |||||.....
 3558 CACCACGGAATATCATGAAAGAGGACGCGCTTAAACAATCTCATCTGCGGAATGT 3607
 |||||.....
 33 rGluProAlaValAlaLeuLeuSerGluAlaGlyLysAsnCysAlaPro 49
 |||||.....
 3608 GGGACCC...TTGGCTTACCTTTGAAGTTAGTGAATCTGTGCT 3654
 |||||.....
 50 ArgGlyLysValArgGlyArgGlyValArgGlnProIleAlaHisTyr 66
 |||||.....
 3655 AGAGAAACCGCTAGCGCGCTTCCGCTTAGCGACGCCATCTGCAGTATAG 3704
 |||||.....
 66 gTPAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluG 83
 |||||.....
 3705 ATGGGACATATGCTAGGCTTGGAGAGCCACGACGAGGATGAGAGAG 3754
 |||||.....
 83 LuanValGlnArgPheGlyLysValArgGlnLeuMetGluLysLeu 99
 |||||.....
 3755 AGATATGGAAAGCTTGGGAGGAGGTGACACAGCTGATCGAAGAGCTG 3804
 |||||.....
 100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 116
 |||||.....
 3805 AGGAAAGCAGTTGAGTCATAGTCTCGCGGCGAGTCAGCAGTATGATCC 3854
 |||||.....
 116 oHISHisAspHisAspGluPheCysLeuMetPro 128
 |||||.....
 3855 TCACCATGACCATCAGATGAGTTTGCCTTATGCC 3891
 |||||.....
 seq_name: gb-pr:AL133348
 seq_documentation_block:
 LOCUS AL133348 40584 bp DNA PRI 06-MAR-2000
 DEFINITION Human DNA sequence from clone RPI-79P11 on chromosome Xq21.32-22.1.
 Contains the gene for a novel protein similar to mouse Bex2
 (brain-expressed x-linked protein 2), ESTs, STSs, GSSs and a
 putative CpG island, complete sequence.
 ACCESSION AL133348
 VERSION AL133348.8 GI:7076386
 KEYWORDS HTG; Bex2; CpG island.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE
 AUTHORS Wilton S
 DIRECT SUBMISSION
 TITLE Submitted (02-MAR-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone

```
CC prevention and diagnosis of cell proliferative, developmental and
CC neurological disorders such as Alzheimer's disease, Huntington's
CC disorders, arteriosclerosis, cancer, atherosclerosis, diabetes,
CC and epilepsy. The CDIFF-4 sequence is homologous to Mus musculus
CC REX-3. This sequence maps to chromosome 1 within the interval from
CC 152.2 to 157.4 centimorgans, to chromosome 3 within the interval from
CC 157.4 to 158.0 centimorgans, and to the X chromosome within the interval
CC from 104.9 to 150.3 centimorgans.
XX
SO Sequence 1364 BP; 411 A; 269 C; 322 G; 362 T; 0 other;

alignment_scores:
    Quality: 170.00      Length: 130
    Ratio: 1.954        Gaps: 6
    Percent Similarity: 66.923   Percent Identity: 34.615

alignment_block:
US-09-327-7500-30 x ACB95548 ..
Align seg 1/1 to: ACB95548 from: 1 to: 1364
1 MetCGLuSerLyAspGln...GLYValIyAasnLeuAanmctGUaaaba 16
|||||.....|.....|.....|.....|.....|.....|
250 ATYGAGTGTCCAAGAGGAACCTAGCGCAACAATCTCACGGGGANAATGC 299
16 phiGInLysLySgLGluLysgLUglULySProGInAapThrILeArgA 33
|||..||..||..||..||..||..||..||..||..||..||..||
30 CCACAAAGAAMAGCAAGAGGGAGGAGGAGGCCCCC....ACGCAAGATC 343
33 TGGLGProALeAlALeULeIleSerGLUaLClyLYLyAenCYaALSPro 49
|||||.....|.....|.....|.....|.....|.....|
344 AAGAAGAATCCGCCCATTTG.....GGAGGGGGTGAGGCCAGAACGCT 387
50 ArGdyGLyArGaTgATgPheATgValARgInProILeALHisTYAf 66
|||||.....|.....|.....|.....|.....|.....|
388 GGAGGAAATATATACGGCGGGGGCGAGCTTAGCGACTTGTCCTAAATTTCG 437
66 gTPraPseUmeGLnArg...ValdYgLUProGIngLyARmetARg 82
|||||.....|.....|.....|.....|.....|.....|
438 ATGGCGCCTACCTATAGGCATATATGACCATAATGAGGAGGA..... 480
82 LUGLUANoVALInatRphedYGLyAspValARgGLImuektGLyLyz 98
|||||.....|.....|.....|.....|.....|.....|
481 ..GATGTGTAGAAAGGTTGTAGGCGCAGATGATGGANATCAAGAGAA 528
59 LeuARgGLuARgCLnLeuSerHIsSerLeuARgALaValISerThASpPr 115
|||||.....|.....|.....|.....|.....|.....|
529 ACTAGGAACACGAGATGAGGCACTATATCGCTTCCAACTCCTCGTAACC 578
115 oProHLshIPasePhLIshAPGLUpHeCYsLeumetPro 128
|||||.....|.....|.....|.....|.....|.....|
579 T.....GACAACCATATGACTTTGCCTTTACT 609

seg_name: /SDS2/gcgdata/geneseq/geneseq/NA2001.DMT:AAF23528

seg_documentation_block:
ID AAF23528 standard; DNA; 700 BP.
AC AAF23528;
AD AAF23528;
DT 22-MAR-2001 (first entry)
DE Mouse NADE DNA.
KW Neurotrophin receptor; p75-NTR; NGF-induced apoptosis;
XX neurogenetic disease; NF-kappaB; ds.
XX
XX Mus sp.
XX WC20007578-A2.
XX 14-DEC-2000
```

Quality: 81.00 Length: 121
Ratio: 1.397 Gaps: 4
Percent Similarity: 47.934 Percent Identity: 23.967
alignment_block:
US-09-327-750d-30 x US-08-459-448A-6 ..
Align seg 1/1 to: US-08-459-448A-6 from: 1 to: 3624
14 GluAsnAspHisGlnLysGlyGluGluLysProGlnAspTh 30
|||||:||||| :|||||:|||||:|||||:|||||:|||||:|||||
1026 GAACACCGACCCACCGCGCCACCAACACGAGCATCAACCGGTGACC 1075
30 rIleArgArgGluProAlaValAlaLeuIleSerGluAlaGlyLysAsnC 47
|||||:||||| :||| :|||||
1076 TCGGCTTCGCCAGCGCCGACGGTGTACCGCACCGAGAGCTACGCCGCGTG 1125
47 YeAla ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..|||
1126 CTGCTGTGGGCATCTACCTCGAGCCCATCCACGGCTGCCACCGTGCG 1175
56 PheArgValArgGlnProIleAlaHisTyrArgTrpAspLeuMetGlnAr 72
|||||:|||||:||||| :||| :||| :||| :||| :||| :|||
1176 CTTCAACTTCACCAACCCCGACACATCAGCCA ..||| ..||| ..|||
72 gValGlyGluProGlnGlyArgMetArgGluGluAsnValGlnArgPheG 89
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1214 GCACGCCCACTACGCCACCCCTACGAGGCC ..||| ..||| ..|||
89 lYcIyAspValArgGlnLeuMetGluLysLeuArgGluArg ..||| ..||| ..|||
1248CGGCTCGAGCTGAGGACACGAGAGCGAGCTGCCGCC 1286
103GlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 117
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1287 CGAGACCGACCGCGCCCACTACGAGAGCTACAGCCACCGCCTGAGCC 1336
117 sHlAspHisHis 121
||| |||||
1337 ACNCGGCATCAT 1349

CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC cDNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNA clones. ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping studies. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.

XX Sequence 532 BP; 151 A; 121 C; 163 G; 91 T; 6 other;

US-09-327-750D-30 x AACF21748 ..

Align seg 1/1 to: AACF21748 from: 1 to: 917

18 GlnLysLysGluGluGlyGluGluGlyProGlnAspThrIleArgArgGl 34
|||||
306 CAGCAAAACAGAGATGGCAGCGCTATCCAGAATGGAGAGAACCG 355
|||||
34 UpToAlaValAlaLeuIleSerGlnAlaGlyLysAANCysAlaPOTArg 51
|||
356 CCTTTG.....GGAGAGGTGAAGGCCACCACCT...G 387
|||||
51 GlylGArgArgPheArgValArgGlnProIleAlaHisTyArgTrp 67
|||||
388 CAGAAATGCAGGGGACAGCGCTCCGCGACTTCGCCCTTAATTTCGTGG 437
|||||
68 AspLeuMetGlnArg..ValGlyGluProGlnGlyArgMetArgGl 83
|||
438 CCNATACCCCATGAGCAGCATCATGATGGATGGTGGA...GATGGAGA 484
|||||
93 uanValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeu 100
|||||
485 TGATATGAGATATTCATCGGAGCATGAGAAATCAGMAAACCTTA 534
|||||
100 RgLuArgGlnLeuSerHisSerLeuArgAlaValSerThrAsp...Pro 115
|||||
535 GGAGCGTCGCAATTCAGGAATTCCTGCCTATCTCTTATGGGGAGCTCTCT 584
|||||
116 ProHisLysAspHisAspGluPheCysLeuMetPro 128
|||||
585 ATNCACCATGACCATCATGATGAAATTTTGCCTTATGCCT 623
|||||

seq_name : /SID2/gcgdata/geneseq/geneseqn/WA2000.DAT: AAC01005
seq_documentation_block:
ID AAC01005 standard; CDNA; 532 BP.
XX AAC01005;
XX
DT 06-OCT-2000 (first entry)
DE Human secreted protein 5' EST, SEQ ID NO: 1003.
DE
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; as.
XX Homo sapiens.
OS
XX
PN EP1033401-A2.
PD
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 990S-0122487.
PR
PA (GEST) GENSET.
PI Dumas Milne Edwards J., Duclert A., Giordano J;
DR WPI; 2000-500381/45.
DR P-PSDB; AAC00999.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1; SEQ ID 1003; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. An ORF has been identified within the

Strd	Orig	zScore	EScore	Len	Documentation
gb.est1.AL022932	+ 683.0	1124.02	1.9e-53	551	! A022932 V6130b28 Beddington m
+ 683.0	1122.79	2.2e-53	642	! AWS36404 G010A11-3 N1A Mouse E	
gb.est2.BG064740	- 680.0	1117.07	4.7e-53	707	! BG064920 H3025D01-3 N1A Mouse E
gb.est2.BG077659	+ 670.0	1102.27	3.1e-52	575	! BG077659 uxa47c11.x1 Soares_NMMA
gb.est2.BG078050	+ 670.0	1100.07	4.1e-52	756	! BG870503 G027915432F1 NCI CGAP_S
gb.est2.BG085459	+ 652.0	1073.78	1.2e-50	459	! BE654459 UT-M-AJ1-aba-f-10-0-UT
gb.est1.AL036394	- 632.0	1039.69	9.5e-49	577	! AWS36394 G0110H01-3 N1A Mouse E
gb.est2.BF067762	+ 591.0	969.71	7.5e-45	785	! BF067762 MY1.000704 Mouse 9-day
gb.est2.BG328077	- 586.0	963.82	1.6e-44	587	! BG328077 uxa47c11.x1 Soares_NMMA
gb.est1.BE9281071	+ 542.50	892.26	1.6e-40	650	! BE9281071 G01086311F1 NCI CGAP_M
gb.est2.BF0680209	+ 537.0	882.33	5.6e-40	583	! BF0680209 MY1.001088 Mouse 9-day
gb.gss.A2936332	- 481.50	791.12	6.7e-35	637	! A2936332 UNO193LO5F Mouse 10kb
gb.est2.BE18183527	- 480.50	788.70	9.1e-35	701	! BE183527 NML-P-FN-by-f-07-0-UNI
gb.est2.D76670	+ 476.50	787.27	1.1e-34	340	! D76670 MUST7D10 mouse embryonal
gb.htc.AK0033429	+ 471.00	776.11	4.6e-34	480	! AK003429 Mus musculus 18 days e
gb.est2.BG696326	+ 466.50	769.94	1.0e-33	512	! BG696326 DRNAEB11 Rat DRG Libr
gb.est1.AV589137	+ 460.00	756.63	5.6e-33	569	! AV589137 Bos taurus br
gb.est1.AW836634	- 456.00	750.34	1.2e-32	519	! AW836634 G0108H08-3 N1A Mouse E
gb.est2.BG655001	+ 453.50	746.25	2.1e-32	547	! BG655001 i0447406.y1 HR85 Islet
gb.est2.BG472167	+ 453.50	745.06	2.5e-32	634	! BG472167 G02513894F1 NIH_MGC_1
gb.est2.BF971303	+ 453.50	744.76	2.6e-32	658	! BF971303 G02273150F1 NIH_MGC_9
gb.est2.BG714823	+ 453.50	744.57	2.6e-32	674	! BG714823 G02677146F1 NIH_MGC_84
gb.est1.AV702285	+ 453.50	744.38	2.7e-32	630	! AV702285 AV702285 ADB Homo sap
gb.est1.AV705808	+ 453.50	744.37	2.7e-32	691	! AV705808 AV705808 ADB Homo sap
gb.est1.AL526209	- 453.50	744.32	2.7e-32	695	! AL526209 AL526209 LIT1.NFL003.NE
gb.est2.BG820179	+ 453.50	744.27	2.7e-32	659	! BG820179 G02782222F1 NIH_MGC_7
gb.est1.BE790774	+ 453.50	743.36	3.1e-32	783	! BE790774 G01581956F1 NCI CGAP_E
gb.est1.AL526247	+ 453.50	743.25	3.1e-32	734	! AL526247 AL526247 LIT1.NFL003.NE
gb.est1.AV702643	+ 450.50	739.52	5.0e-32	683	! AV702643 ADB Homo sap1
gb.est1.BE314909	+ 450.00	739.82	4.8e-32	594	! BE314909 G01144007F1 NIH_MGC_9
gb.est1.AL520931	- 450.00	738.31	5.8e-32	717	! AL520931 AL520931 LIT1.NFL0004.NE
gb.est2.BE889313	+ 450.00	737.71	6.3e-32	772	! BE889313 G01513201F1 NIH_MGC_71
gb.est2.BG709503	+ 450.00	737.32	6.6e-32	810	! BG709503 G02674781F1 NIH_MGC_96
gb.est2.BG707398	+ 450.00	737.08	6.8e-32	835	! BG707398 G02672810F1 NIH_MGC_96
gb.est2.BG3285825	+ 450.00	736.07	7.8e-32	946	! BG3285825 G02380917F1 NIH_MGC_93
gb.est2.BF038787	+ 450.00	735.29	8.6e-32	1042	! BF038787 G01462119F1 NIH_MGC_93
gb.est1.AI929703	+ 446.50	734.31	9.8e-32	576	! AI929703 au63074.y1 Schneider
gb.est2.BG715659	+ 446.50	732.35	1.3e-31	734	! BG715659 G02676924F1 NIH_MGC_96
gb.est2.BG708231	+ 446.00	735.57	8.3e-32	445	! BG708231 H3025D01-5 N1A Mouse E
gb.est1.BE03782324	+ 444.50	731.73	1.4e-31	537	! BE037823 130127-MARC 1P1G Sus s
gb.est1.BE031524	+ 444.50	731.73	1.4e-31	547	! BE031524 130127-MARC 1P1G Sus s

ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-Sal3 (include SalI sequence). The cDNAs were purified by phenol/chloroform and separated from free linkers by Centricon 100. Then, cDNAs were amplified by long-range high fidelity PCR using Takara's Ex Taq polymerase. Then, the cDNAs were purified by phenol/chloroform and by Centricon 100. The cDNAs were digested with SalI and NotI enzymes. Then, the cDNAs were size selected by Gibco's Size Fractionation Column. The cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by chemical method. The library was constructed by Xiaohong Wang and Minoru S. H. Ko.

BASE COUNT 145 a 182 c 133 g 182 t
ORIGIN

alignment_scores:
Quality: 683.00 Length: 128
Ratio: 5.378 Gaps: 0
Percent Similarity: 99.219 Percent Identity: 99.219

alignment_block:
US-09-327-750D-30 x AW536404/rev

Align seg 1/1 to reverse of: AW536404 from: 1 to: 642

```

1 MetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspHi 17
|||||
606 ATGGAGTCCAAAGATCAAGGCGTGAAAAATCTCAACATGGAGATGACCA 557
|||||
17 sGlnLysLysGluGluLysGluLysGluLysProGlnAspThrIleArg 34
|||||
556 TCAGAAAAAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 507
|||||
34 luProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50
|||||
506 AGCCACTGTGCCCTCGACCTCCGAGGCTGGCAAAACTGTGCACCTAGA 457
|||||
51 GlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrArgTr 67
|||||
456 GGAGGTCGCGAGCGGTTCGGGTTCGGCAGGCCCATCGCTCACTATAGTG 407
|||||
67 pAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluAla 84
|||||
406 GGACCTGATGCAGAGGGTTGGGAGCCCCCAGGAGGATGAGAGAGAGA 357
|||||
84 snValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuArg 100
|||||
356 ACGTACAGAGGTTTGGGGGTGATGTGAGACAGCTCATGGAGAGCTGAGG 307
|||||
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
|||||
306 GAAGGCACCTGAGCCACAGCTCGGGGGGTAGCAGCTGACCTGACCCGCTCA 257
|||||
117 sHisAspHisHisAspGluPheCysLeuMetPro 128
|||||
256 TCATGACCACCATGATGAGTTTTCCTCCTATGCC 223
|||||
seq_name: gb_est2:BG064920

```

seq_documentation_block:

LOCUS BG064920 707 bp mRNA EST 26-JAN-2001
DEFINITION H3025D01-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
ACCESSION BG064920
VERSION BG064920.1 GI:12547483
KEYWORDS EST
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 707)

```

34 luProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50
|||||
242 AGCCACTGTGCCCTCGACCTCCGAGGCTGGCAAAACTGTGGCCTAGA 291
|||||
51 GlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrArgTr 67
|||||
292 GGAGGTCGCGAGCGGTTCGGGTTCGGCAGGCCCATCGCTCACTATAGTG 341
|||||
67 pAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluAla 84
|||||
342 GGACCTGATGCAGAGGGTTGGGAGCCCCCAGGAGGATGAGAGAGAGA 391
|||||
84 snValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuArg 100
|||||
392 ACGTACAGAGGTTTGGGGGTGATGTGAGACAGCTCATGGAGAGCTGAGG 441
|||||
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
|||||
442 GAAGGCACCTGAGCCACAGCTCGGGGGGTAGCAGCTGACCTGACCCGCTCA 491
|||||
117 sHisAspHisHisAspGluPheCysLeuMetPro 128
|||||
492 TCATGACCACCATGATGAGTTTTCCTCCTATGCC 525
|||||
seq_name: gb_est1:AW536404

```

seq_documentation_block:

LOCUS AW536404 642 bp mRNA EST 31-AUG-2000
DEFINITION G0104A11-3 NIA Mouse E7.5 Embryonic Portion cDNA Library Mus
musculus cDNA clone G0104A11 3', mRNA sequence.
ACCESSION AW536404
VERSION AW536404.1 GI:7178821
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 642)
Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac
M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H., Wood,W.H.
III, Becker,K.G. and Ko,M.S.H.
Genome-wide expression profiling of mid-gestation placenta and
embryo using a 15,000 mouse developmental cDNA microarray
proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
20381348
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: G0104 row: A column: 11
Seq primer: -21M13 Forward
High quality sequence stop: 642
POLYA-yes.
Location/Qualifiers
1 . 642
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G0104A11"
/clone_lib="NIA Mouse E7.5 Embryonic Portion cDNA Library"
/sex="unknown"
/dev_stage="7.5dpc Embryo"
/lab_host="DH10B"
/note="vector: pSPORT1 (Gibco/BRL Life Technology);
Site 1: SalI; Site 2: NotI; Total RNAs were extracted from
6 Embryo. The double-stranded cDNA was synthesized by
Gibco's kit with an Oligo(dT) primer (NotI primer-adaptor
from Gibco/BRL)
[5'-pGACTAGTTCTAGTCGCGAGCGGCCGCTTTTCTTTT-3']
from 0.51ug of mRNA. The double-stranded cDNAs were
treated with T4 DNA polymerase and purified by

FEATURES
source

AUTHORS Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A., Tanaka, T.S., Carter, M.G. and KO, M.S.H.
TITLE Verification and initial annotation of NIA mouse 15K cDNA clone set
JOURNAL Unpublished (2001)
COMMENT Other_ESTs: H3025D01-5
 Contact: George J. Kargul
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdnaelg@nig.nih.gov
 This clone set has been freely distributed to the community. Please visit <http://lgsun.grc.nia.nih.gov/cDNA/15k.html> for details.
 Plate: H3025 row: D column: 01
 Seq primer: -21M13 forward
 High quality sequence stop: 707
 POLYA=Yes.

FEATURES

source
 1. 707
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="H3025D01"
 /clone.lib="NIA Mouse 15K cDNA Clone Set"
 /sex="Clones arrayed from a variety of cDNA libraries"
 /dev_stage="Clones arrayed from a variety of cDNA libraries"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with oligo(dT) -Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."
 BASE COUNT 154 a 209 c 146 g 198 t
 ORIGIN

alignment_scores:
 Quality: 680.00 Length: 128
 Ratio: 5.354 Gaps: 0
 Percent Similarity: 99.219 Percent Identity: 98.438
alignment_block:
 US-09-327-750d-30 x BG064920/rev ..
 Align seq 1/1 to reverse of: BG064920 from: 1 to: 707
 1 MetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspH1 17
 606 ATGGAGTCCAAAGATCAAGGCGTGAATAATCTCAACATGGAGATGACCA 557
 17 sGlnLysLysGluGluLysGluGluProGlnAspThrIleArgArg 34
 556 TCAGAAAGAGGAGAGAGGAGAGAGAAACCAAGATACCATCAGAGGG 507
 34 luProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50
 506 AGCCAGCTGTGGCCCTGACCTCCGAGGCTGGCAAAACTGTGCACCTAGA 457
 51 GlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrArgTr 67

|||||
 456 GGAGTGGCAGGCGGTTCGGGTTCGGCAGCCCATCGCTCACTATAGATG 407
 67 PAspLeuMetGlnArgValGlyValProGlnGlyArgMetArgGluGluA 84
 406 GGAGCTGATCAGAGGTTGGGGAGCCCGAGGAGGATGAGAGAGGAGA 357
 84 snValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuArg 100
 356 ACGTACAGAGGTTTGGGGGTGATGTGAGACAGCTCATGGAGAAGCTGAGG 307
 101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH1 117
 306 GAAAGGAGCTGAGCCACAGCGTGGCGGCTAGCAGCTGACCCGCCCTCA 257
 117 SHISAspHisHisAspGluPheCysLeuMetPro 128
 256 TCATGACCACCATGATGAGTTTTCCTCATGCC 223
 seq_name: gb_esf2:BG277659
 seq_documentation_block:
 LOCUS BG277659 575 bp mRNA EST 21-FEB-2001
 DEFINITION ux47c11.y1 Soares_NMAX_maxillary_process Mus musculus cDNA clone
 IMAGE:3513237 5' similar to TR:Q9RJ2 Q9RJ2 BEX1 PROTEIN. ; mRNA
 SEQUENCE.
 ACCESSION BG277659
 VERSION BG277659.1 GI:13073183
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 1 (bases 1 to 575)
 NC1-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
 Tumor Gene Index
 Unpublished (1997)
 JOURNAL
 COMMENT Other_ESTs: ux47c11.x1
 Contact: Robert Strausberg, Ph.D.
 Email: cgapb-r@mail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (infoimage.llnl.gov) for further information.
 MGI:11393813
 Seq primer: -40RP from Gibco
 High quality sequence stop: 462.
 Location/Qualifiers
 1. 575
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:3513237"
 /clone.lib="Soares_NMAX_maxillary_process"
 /tissue_type="maxillary process"
 /lab_host="DH10B (phage-resistant)"
 /note="Vector: p7T3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTTCACATCTGAAGTGGAGGCGCGCCCTTTTCTTTTCTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified p7T3 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 156 a 136 c 201 g 81 t 1 others
 ORIGIN
 alignment_scores:
 Quality: 670.00 Length: 128
 Ratio: 5.317 Gaps: 0
 Percent Similarity: 98.438 Percent Identity: 97.656
 alignment_block:

US-09-327-750D-30 x BG277659

Align seg 1/1 to: BG277659 from: 1 to: 575

```
1 MetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspH1 17
|||||
192 ATGGAGTCCAAAGATCAAGCGGTGAAGAAATCTCAACATGGAGATGACCA 241
|||||
17 sGlnLysLysGluGluLysGluGluLysProGlnAspThrIleArgArg 34
|||||
242 TCAGAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 291
|||||
34 luProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50
|||||
292 AGCCAGCTGTGGCCCTGACCTCCGAGGCTGGCAAAACTGTGCACCTAGA 341
|||||
51 GlyGlyArgArgArgPheArgValArgGlnProIleAlaHisTyrArgTr 67
|||||
342 GGAGTGCAGCGGTTCCGGGTTCCGAGCCCATCGCTCACTATAGATG 391
|||||
67 pasLeuMetClnArgValGlyGluProGlnGlyArgMetArgGluGlu 84
|||||
392 GGACTGTATGAGAGGTTGGGAGGCCCGGAGGAGGATGAGAGAGGAGA 441
|||||
84 snValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuArg 100
|||||
442 ACGTACAGAGGTTGGGAGGTTGATGTGAGACAGCTCATGGAGAGCTGAGG 491
|||||
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH1 117
|||||
492 GAAAGGAGCTGTAGCCACACCTCGCGGGGTTAGCACTGACCCGCCCTCA 541
|||||
117 sHisAspHisAspGluPheCysLeuMetPro 128
|||||
542 TCATGACCACTATGATGATGTTGCCTCATGCCCC 575
|||||
```

seq_name: gb_est2:BG870503

```
seq_documentation_block: 756 bp mRNA EST 29-MAY-2001
LOCUS BG870503 602791422F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4922644 5',
DEFINITION mRNA sequence.
ACCESSION BG870503
VERSION BG870503.1 GI:14221043
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10842 row: 1 column: 05
High quality sequence stop: 756.
Location/Qualifiers
```

```
FEATURES
source
1..756
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4922644"
/lab_host="NCI_CGAP_SG2"
/note="organ: salivary gland; Vector: pcwv-sport6; Site:1"
```

NotI; Site:2; Sali; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library.

BASE COUNT 207 a 164 c 230 g 155 t
ORIGIN

alignment_scores:
Quality: 670.00 Length: 129
Ratio: 5.276 Gaps: 1
Percent Similarity: 98.450 Percent Identity: 98.450

alignment_block:

US-09-327-750D-30 x BG870503

Align seg 1/1 to: BG870503 from: 1 to: 756

```
1 MetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspH1 17
|||||
165 ATGGAGTCCAAAGATCAAGCGGTGAAGAAATCTCAACATGGAGATGACCA 214
|||||
17 sGlnLysLysGluGluLysGluGluLysProGlnAspThrIleArgArg 34
|||||
215 TCAGAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 264
|||||
34 luProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50
|||||
265 AGCCAGCTGTGGCCCTGACCTCCGAGGCTGGCAAAACTGTGCACCTAGA 314
|||||
51 GlyGlyArgArgArgPheArgValArgGlnProIleAlaHisTyrArgTr 67
|||||
315 GGAGTGCAGCGGTTCCGGGTTCCGAGCCCATCGCTCACTATAGATG 364
|||||
67 pasLeuMetClnArgValGlyGluProGlnGlyArgMetArgGluGlu 83
|||||
365 GGACTGTATGAGAGGTTGGGAGGCCCGGAGGAGGATGAGAGAGGAG 414
|||||
84 AsnValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuAr 100
|||||
415 AACGTACAGAGGTTGGGAGGTTGATGTGAGACAGCTCATGGAGAGCTGAG 464
|||||
100 gGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 117
|||||
465 GGAAGGAGGAGCTGAGCCACACCTCGCGGGGTTAGCACTGACCCGCCCTC 514
|||||
117 sHisAspHisAspGluPheCysLeuMetPro 128
|||||
515 ATCATGACCACTATGATGATGTTGCCTCATGCCCC 549
|||||
```

seq_name: gb_est1:BE654459

```
seq_documentation_block: 499 bp mRNA EST 06-SEP-2000
LOCUS BE654459 UI-M-AJ1-aha-f-10-0-UI.f1 NIH_BMAP_MOB_N Mus musculus cDNA clone
DEFINITION UI-M-AJ1-aha-f-10-0-UI 5', mRNA sequence.
ACCESSION BE654459
VERSION BE654459.1 GI:9980372
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 499)
Normalizing and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
```

Fax: 301 443 9890

Email: mestr@mail.nih.gov

CDNA Library Preparation: M.B. Soares Lab Clóne distribution: Researchers may obtain BMAP CDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements

Seq primer: M13 Reverse

FEATURES

Location/Qualifiers

source

1..499

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UI-M-AJ1-aba-f-10-0-UI"

/clone_lib="NIH_BMAP_MOB_N"

/dev_stage="27-32 days"

/lab_host="DH10B (Life Technologies)"

/note="vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_MOB_N library is a normalized library constructed from mouse olfactory bulbs. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories."

146 a 108 c 179 g 66 t

BASE COUNT

ORIGIN

alignment_scores:

Quality: 652.00 Length: 123

Ratio: 5.344 Gaps: 0

Percent Similarity: 99.187 Percent Identity: 99.187

alignment_block:

US-09-327-750D-30 x BE654459

Align seg 1/1 to: BE654459 from: 1 to: 499

```

1 MetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspH1 17
131 ATGGAGTCCAAAGATCAAGCGCTGAAAAAATCTCAACATGGAGAATGACCA 180
17 sGlnLysLysGluGluLysGluLysProGlnAspThrIleArgArg 34
181 TCAGAAAAGGAGGAGAGAGAGAAAGCCACAGATACCATCAGAGGG 230
34 luProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50
231 AGCCAGCTGTGGCCCTGACCTCCGAGGCTGGCAAAACTGTGCACCTAGA 280
51 GlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrArgTr 67
281 GGAGGTGCGAGCGGTTCGGGTTCGGAGCCCATCGCTCACTATAGATG 330
67 pAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluAla 84
331 GGACCTGATGCGAGAGGTGGGGAGCCCGGAGGAGGATGAGAGAGGAGA 380
84 snValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuArg 100
381 ACGTACAGAGGTTGGGGGTGATGTGACACAGCTCATGGAGAACTGAGG 430
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH1 117
431 GAAAGGCAGCTGAGCCACAGCTCGGGCGGTAGCACTGACCGCGCTCA 480
117 sHisAspHisHisAspGlu 123
481 TCATGACCACTCATGATGAG 499

```

seq_name: gb_estl:AW536974

seq_documentation_block:

LOCUS AW536974

DEFINITION G0110H10-3 N1A Mouse E7.5 Embryonic Portion cDNA Library Mus

musculus cDNA clone G0110H10 3', mRNA sequence.

ACCESSION AW536974

VERSION AW536974.1 GI:7179391

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus

REFERENCE 1 (bases 1 to 577)

AUTHORS Tanaka, T.S., Jaredat, S.A., Lim, M.K., Kargul, G.J., Wang, X., Grahovac

M.J., Pantano, S., Sano, Y., Piao, Y., Nagaraja, R., Doi, H., Wood, W.H., III, Becker, K.G. and Ko, M.S.H.

TITLE Genome-wide expression profiling of mid-gestation placenta and

embryo using a 15,000 mouse developmental cDNA microarray

Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)

MEDLINE 20381348

COMMENT Contact: George J. Kargul

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@gsun.grc.nia.nih.gov

Plate: G0110 row: H column: 10

Seq primer: -21M13 Forward

High quality sequence stop: 577

POLYA=Yes.

FEATURES

Location/Qualifiers

1..577

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_lib="G0110H10"

/sex="unknown"

/dev_stage="7.5dpc Embryo"

/lab_host="DH10B"

/note="vector: pSPORT1 (Gibco/BRL Life Technology);

Site_1: SalI; Site_2: NotI; Total RNAs were extracted from

6 Embryo. The double-stranded cDNA was synthesized by

Gibco's kit with an Oligo(dT) primer (NotI primer-adaptor

from GibcoBRL)

[5'-pGACTAGTCTCATGCGAGCGCGCCCTTTTTTTTTTTT-3']

from 0.5µg of mRNA. The double-stranded cDNAs were

treated with T4 DNA polymerase and purified by

ethanol-precipitation. The cDNAs were ligated to

Lone-linker LL-Sal3 (include SalI sequence). The cDNAs

were purified by phenol/chloroform and separated from

free linkers by Centricon 100. Then, cDNAs were amplified

by long-range high fidelity PCR using Takara's Ex Taq

polymerase. Then, the cDNAs were purified by

phenol/chloroform and by Centricon 100. The cDNAs were

digested with SalI and NotI enzymes. Then, the cDNAs were

size selected by Gibco's Size Fractionation Column. The

cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid

vector. The DH10B E. coli host was transformed with the

ligation mixture by chemical method. The library was

constructed by Xiaohong Wang and Minoru S. H. Ko."

BASE COUNT 137 a 160 c 126 g 154 t

ORIGIN

alignment_scores:

Quality: 632.00 Length: 118

Ratio: 5.402 Gaps: 0

Percent Similarity: 99.153 Percent Identity: 99.153

alignment_block:

US-09-327-750D-30 x AW536974/rev

Align seg 1/1 to reverse of: AWS36974 from: 1 to: 577

```

11 LeuAsnMetGluAsnAspHisGlnLysGluGluLysGluLysGluLysPr 27
|||||
576 CTCACATGGAGATGACCATCAGAAAAGGAGGAGAGGAGAGGAGAGGCC 527
|||||
27 oglnAspThrIleArgArgGluProAlaValAlaLeuIleSerGluAlaG 44
|||||
526 ACAAGATACCATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 477
|||||
44 lYlYsAsnCysAlaProArgGlyGlyArgArgPheArgValArgGln 60
|||||
476 GCANAACTGTGCACCTAGAGAGGTCGACAGGCGTTCGGGTTCGGCAG 427
|||||
61 ProIleAlaHisTyrArgTTrpAspLeuMetGlnArgValGlyGluProG 77
|||||
426 CCCATCGCTCACTATAGATGGACCTGATCAGAGGGTTGGGAGGCCCA 377
|||||
77 nGlyArgMetArgGluGluAsnValGlnArgPheGlyGlyAspValArgG 94
|||||
376 GGAAGAGTACAGAGGAGACGTACAGAGGTTGGGGTGATGTGAGAC 327
|||||
94 lNleuMetGluLysLeuArgGluGlnLeuSerHisSerLeuArgAla 110
|||||
326 AGCTCATGAGAAAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 277
|||||
111 ValSerThrAspProProHisHisAspHisHisAspClnPheCysLeuMe 127
|||||
276 GTTACACTGACCCGCTCATCATGACCATGATGATGATGATGATGATG 227
|||||
127 tPro 128
|||||
226 GCCC 223

```

seq_name: gb_est2:BF607762

```

seq_documentation_block: 785 bp mRNA EST 01-APR-2001
LOCUS BF607762
DEFINITION MT1_000704 Mouse 9-day fetus cDNA library ICRFp522 Mus musculus
cDNA clone ICRFp522A1947 5', mRNA sequence.
ACCESSION BF607762
VERSION BF607762.1 GI:13504254
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Yahyawi, M., Hennig, S., Neidhardt, L., Radelof, U., Hermann, B.G.,
Lehrach, H. and O'Brien, J.
TITLE Detection of a high number of novel genes in a 9-day mouse embryo
cDNA library normalised by oligonucleotide fingerprinting
JOURNAL Unpublished (2001)
COMMENT Contact: Hennig S
Laboratory 123, Dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr. 63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1612
Fax: +49 30 8413 1380
Email: hennig@molgen.mpg.de
ESTs are made from clones being representatives of clone clusters.
Clone clusters were calculated from oligonucleotide fingerprints.
PCR Primers
FORWARD: 5'-GAGTATTCCAGAGTAGTGA-3'
BACKWARD: 5'-TAATGACCTACTATAGG-3'
Seq primer: 5'-ATTAGTGACATAG-3'
High quality sequence stop: 785.
location/Qualifiers
source 1.785
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="ICRFp522A1947"
/clone_lib="Mouse 9-day fetus cDNA library ICRFp522"

```

FEATURES

```

/tissue_type="whole embryo"
/dev_stage="embryonic 9-day"
/lab_host="E.coli, XL1 blue"
/notes="Vector: pSVSPORT1; Site.1: NotI; Site.2: SalI;
Library preparation by oligo.dT priming of RNA. Clones can
be ordered from the Resource Center in Berlin,
http://www.rzpd.de."
BASE COUNT 190 a 188 c 233 g 153 t 21 others
ORIGIN

```

```

alignment_scores:
Quality: 591.00 Length: 128
Ratio: 4.844 Gaps: 4
Percent Similarity: 95.312 Percent Identity: 93.750
alignment_block:
US-09-327-750d-30 x BF607762
Align seg 1/1 to: BF607762 from: 1 to: 785

```

```

1 MetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspHi 17
|||||
259 ATGGAGTCCAAAGATCAAGGCGTGAATAATCTCAACATGGAGAATGACCA 308
|||||
17 sGlnLysLysGluGluLysGluLysProGlnAspThrIleArgArg 34
|||||
309 TCNAGAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 358
|||||
34 lNleuProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50
|||||
359 AGCCAGCTGTGCCCTGACCTCCGAGGCTGGCAAAACTGTGCACCTAGA 408
|||||
51 GlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrArgTr 67
|||||
409 GGAGGTGCGAGCGGTTCGGGTTCGGCAGCCCATCGCTCACTATAGATG 458
|||||
67 pAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluAla 84
|||||
459 GGACCTGTATGTCAGAGGGTTGGGAGGCCCGAGGAGGATGAGAGAGGAGA 508
|||||
84 sNValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuAr 100
|||||
509 ACGTACAGAGGTTTGGGGTGATGTGACACAGETCATGGGGGGAACCTGAG 558
|||||
100 gLluArgGlnLeuSerHisSerLeuArgAlaValSerThrAsp.ProPro 116
|||||
559 GGAAGGACCTGAAACCCACAGCCTGCGGGCGGTAGCAGCTGACCCCGCCT 608
|||||
117 His.HisAspHisHis.AspGluPhe 124
|||||
609 CATTCATGACCACTGATGATGATGATGATGATGATGATGATGATGAT 634
|||||
seq_name: gb_est2:BG228077

```

seq_documentation_block:

```

LOCUS BG228077 587 bp mRNA EST 08-FEB-2001
DEFINITION BX47C11.X1 Soares.NMAMAX_maxillary process Mus musculus cDNA clone
IMAGE:3513237 3' similar to TR:09R1J2 09R1J2 BEX1 PROTEIN. ; mRNA
sequence.
ACCESSION BG228077
VERSION BG228077.1 GI:12715592
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 587)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.

```


(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b/AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT 131 a 197 c 132 g 176 t 1 others
ORIGIN

alignment_scores:
Quality: 481.50 Length: 117
Ratio: 4.500 Gaps: 1
Percent Similarity: 91.453 Percent Identity: 84.615

alignment_block:
US-09-327-750D-30 x A2936393/rev ..

Align seg 1/1 to reverse of: A2936393 from: 1 to: 637

1 MetGluSerLys...AspGlnGlyValLysAsnLeuAsnMetGluAsnAs 16
|||||
382 ATGGAGTCCAAAGTGGAAACAGCGCTGAAANCTCAACATGAGGATGA 333
16 phisGlnLysLysGluGluLysGluLysProGlnAspThrIleArgA 33
|||||
332 CCATCAGGAAAGAGGAGGAAAGGAAAGGAAAGGAAAGGAAAGGAA 283
33 rgGluProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaPro 49
|||||
282 GGGATCCGATGTGGCCCTGCTTCGAGCTGGAGCTACTAGTGCCT 233
50 ArgGlyArgArgArgPheArgValArgGlnProIleAlaHisTyrAr 66
|||||
232 AGAGGAGTTCGAGCGGCTTCGGGTTCCGAGCTGGAGCTACTAGTGC 183
66 gtrpAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGlu 83
182 ATGGACCTGATGCTAGAGGTTCGGGAGCCCGGAGGAGGATGAGAGG 133
83 luAsnValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeu 99
|||||
132 AGAAGCTACAGAGGTTCGGGATGATGTGAGACAGCTCATGGAGAGCT 83
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 116
|||||
82 GG.GAAGGCACCTAGCCACAGCTTCGGGCGGTTCAGCTAGCCGCC 34
116 o 116
33 T 33

seq_name: gb_est2-B1183527

seq_documentation_block:
LOCUS B1183527 701 bp mRNA 10-JUL-2001
DEFINITION UNL-P-FN-by-f-07-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
UNL-P-FN-by-f-07-0-UNL 3', mRNA sequence.
ACCESSION B1183527
VERSION B1183527
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 701)
Caetano, A.R., Johnson, R.K. and Pomp, D.
Generation and sequence characterization of a normalized cDNA library from swine ovarian follicles
Unpublished (2001)
Contact: Pomp, D
Department of Animal Science
University of Nebraska, Lincoln
Lincoln, NE 68583-0908, USA
Tel: 402 472 6416
Fax: 402 472 6362
Email: dpomp@unl.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized porcine ovarian follicles library
Seq primer: M13 -29
POLYA=Yes.

FEATURES
source
Location/Qualifiers
1..701
/organism="Sus scrofa"
/strain="University of Nebraska, Lincoln Swine Selection Lines"
/db_xref="taxon:9823"
/clone="UNL-P-FN-by-f-07-0-UNL"
/clone_lib="UNL-P-FN"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN library is a normalized library representing porcine ovarian follicles, ranging from 2.0 to 10.0 mm in diameter, collected during 7 days of the follicular phase of the pig estrous cycle. This library was derived from the library UNL-P-F2. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.
TAG_LIB=UNL-P-FN
TAG_TISSUE=porcine ovarian follicles
TAG_SEQ=CACACT"

BASE COUNT 173 a 190 c 119 g 218 t 1 others
ORIGIN

alignment_scores:
Quality: 480.50 Length: 130
Ratio: 4.215 Gaps: 2
Percent Similarity: 87.692 Percent Identity: 71.538

alignment_block:
US-09-327-750D-30 x B1183527/rev ..

Align seg 1/1 to reverse of: B1183527 from: 1 to: 701

1 MetGluSerLysAsp.....GlnGlyValLysAsnLeuAsnMetGluAs 15
|||||
653 ATGGCGTCCAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 604
15 nAspHisGlnLysLysGluGluLysGluLysProGlnAspThrIleA 32
|||||
603 TACCACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 554
32 rgArgGluProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAla 48
|||||
553 AAGGAGAGCCT...TTGGCCCTCCCTGTGGAGCTGGTGGTGGTGGT 507
49 ProArgGlyArgArgArgPheArgValArgGlnProIleAlaHisTyr 65

```

|||||.....|
506 CCTAGAGGAATCGTAGCGGTTCCGTGTGAGGAGCCCTATCTCGCAGTA 457
|||||.....|
65 rArgTTPasPLeuMetGlnArgValGlyGluProGlnGlyArgMetArg 82
|||||.....|
456 TAGATGGATATGACGAGAGGCTTGAGAGCCACAGGAGATGAGAG 407
|||||.....|
82 luGluAsnValGlnArgPheGlyClyAspValArgGlnLeuMetGluLys 98
|||||.....|
406 AAGAGAATATGAAAGGATTGGGAGGAGGTGAGGCTGCTGATGGAAG 357
|||||.....|
99 LeuArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPr 115
|||||.....|
356 CTGAGGGAAGAGCAGTTGATCATAGTCTTCGGGCAGTTAGCACATGCC 307
|||||.....|
115 pProHisAspHisHisAspGluPheCysLeuMetPro 128
|||||.....|
306 CCTCACCATGACCATCAGCATGATGATTTGCCTTATGCCT 267

```

seq_name: gb_est2:D76670

seq_documentation_block: 370 bp mRNA EST 07-OCT-1996
LOCUS D76670
DEFINITION M0577D10 mouse embryonal carcinoma cell line F9 Mus musculus cDNA
clone 77D10, mRNA sequence.

ACCESSION D76670
VERSION D76670
KEYWORDS EST
SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 370)
AUTHORS Jishiguchi, S., Sakuma, R., Nomura, M., Zou, Z., Jearnaislavong, J.,
Joh, T., Yasunaga, T. and Shimada, K.
TITLE A catalogue of genes in mouse embryonal carcinoma F9 cells
identified with expressed sequence tags
J. Biochem. (Tokyo) 119 (4), 749-767 (1996)
96337530

JOURNAL

Medline
COMMENT Contact: Kazunori Shimada
Department of Medical Genetics, Division of Molecular Biomedicine
Research Institute for Microbial Diseases, Osaka University
3-1, Yamadaoka, Suita, Osaka, 565, Japan
Tel: 06-875-8325
Fax: 06-875-8326.

FEATURES

source
Location/Qualifiers
1..370
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="77D10"
/clone_lib="mouse embryonal carcinoma cell line F9"
BASE COUNT 110 a 72 c 124 g 53 t 1 others
ORIGIN

alignment_scores:

Quality: 476.50 Length: 96
Ratio: 5.124 Gaps: 1
Percent similarity: 96.875 Percent identity: 96.875

alignment_block:

US-09-327-750d-30 x D76670 ..
Align seg 1/1 to: D76670 from: 1 to: 370

```

1 MetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspH 17
|||||.....|
68 ATGAGTCCAAAGATCAAGCGGTGAAATCTCAATGAGATGACCA 117
|||||.....|
17 scLysLysGluGluLysGluLysProGlnAspThrLeArgArg 34
|||||.....|
118 TCAGAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 167

```

```

34 luProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50
|||||.....|
168 ACCAGAGCTGTGCCCTGATCTCCGAGGCTGGCAAAACTGTGGCCTAGA 217
|||||.....|
51 GlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrArgTr 67
|||||.....|
218 GGAGGTCCGAGCGGTTCCGGGTTCCGAGCCCATCGCTCACTATAGATG 267
|||||.....|
67 pAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluGlu 84
|||||.....|
268 GGACCTGATGCAGAGGTTGGGAGCCCGAGGAGGAGGATGAGAGGAGA 317
|||||.....|
84 snValGlnArgPheGlyCly...AspValArgGlnLeu 95
|||||.....|
318 ACGTACAGAGGTTTNGGGGTTATTGTTAGCAACTT 355

```

seq_name: gb_htc:AK003429

seq_documentation_block:

LOCUS AK003429 480 bp mRNA HTC 05-JUL-2001
DEFINITION Mus musculus 18 days embryo cDNA, RIKEN full-length enriched
library, clone:1110004J10, full insert sequence.

ACCESSION AK003429

VERSION AK003429.1 GI:12834096

KEYWORDS CAP trapper.

SOURCE Mus musculus (strain:C57BL/6J) 18 days embryo cDNA to mRNA,
clone:1110004J10.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 480)
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

AUTHORS

2 (bases 1 to 480)
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome research. 10 (10), 1617-1630 (2000)
20499374

JOURNAL

Medline
PUBMED 11042159

REFERENCE

3 (bases 1 to 480)
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome research. 10 (11), 1757-1771 (2000)

AUTHORS

4 (bases 1 to 480)
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

JOURNAL

Medline
PUBMED 11076861

REFERENCE

5 (bases 1 to 480)
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Araki, A.,
Arakawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M.,
Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F.,
Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J.,
Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T.,
Miyazaki, A., Nishi, K., Nomura, K., Numata, R., Ohno, M., Okazaki, Y.,
Okido, T., Owa, C., Saito, H., Saito, K., Sakai, K., Sano, H.,
Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,

Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE
JOURNAL

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-2 Suenihiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome.res@gs.c.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216).

fax: 01 43 303 9210)
please visit our web site (<http://genome.gsc.riken.go.jp/>) for
further details.

RNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5', GAGAGAGAGCGCGCACTCGAGTGTGTTTATTTTATN 3'], cDNA was prepared by using triose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0. Second strand cDNA was prepared with the primer adapter of sequence[5', GAGAGAGAGAGATTCCTAAGCTCAATTAATTAATTAACCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end; SstI, 3' end;

FEATURES

```

1. .480
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGD:MGI:1328321"
/db_xref="MGD:MGI:1900507"
/clone="1110004J10"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="18 days embryo"
<1. .261
/note="putative"
/codon_start=1
/protein_id="BAB22786.1"
/db_xref="GI:12834097"
/translation="SGNKACAPGRRRRRRVPRIAHYEWDLQWVGQPGQRMREINVQ
PEGEVCLMEKTEPEEOLSHSTADPEHHDEECLAMP"

```

polyA_signal

polyA_site

BASE COUNT	117 a	103 c	132 g	128 t
ORIGIN	/note="putative"			

alignment scores:

Quality:	471.00	Length:	86
Ratio:	5.477	Gaps:	0
Similarity:	100.000	Percent Identity:	98.837

Ratio:	5.477	Gaps:	0
--------	-------	-------	---

Similarity: 100.000 Percent Identity: 98.837

alignment_block:

US-09-327-750D-30 x AK003429

Align seq 1/1 to: AK003429 from: 1 to: 480

43 AlaGlyIysAsnCysAlaProArgGlyGlyArgArgPheArgValar 59
:::|||||
1 TCTGGCAAAACTGTGCACCTAGAGGAGTTCGAGCGGTTCCGGSTTCG 50

1 TCTGGCAAAACTGTGCACCTAGAGGAGGTCGCAGGCGGTTCGGGTTGG 50

59 aGlnProIleAlaHisTyrArgATrpAspLeuMetGlnArgValGlyGly 76

[illegible]

51 GCAGCCCATCGCTCACCTATAGATGGGACCCTGATGCAGAGGGTTGGGGAGC 10

76 roGlnGlvArOMeTArGlllGlllAsnValGlnArpHeGlvGlvAsnVal 92

[illegible]

151 AGCAGCTCATGGAGAACTCAGGGAAGCAGCTGAGCCACAGCTCG 200
109 gAlaValSerThrAspProHisHisAspHisHisAspGluPheCysL 126
201 GCGCGTTAGCACTGACCGCGCTCATCATGACCACCATGATGAGTTTGCC 250
126 euMetPro 128
251 TCATGCC 258

109 qAlavalSerThrAspProHisHisAspHisHisAspGluPheCysL 126

[illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

126 euMetPro 128

OM of: US-09-327-750D-30 to: Issued_Patents_NA:* out_format : pfs
Date: Mar 11, 2002 3:50 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL-frame-pzn.model -DEV-xlh
-Q/cgn2_1/USPTO.spool/US09327750/runat_11032002_101154_20340/app_query.fasta_1.1472
-DB-Issued_Patents_NA -OPT=fastcap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=bloms62
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTENT=pfs
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09327750@cgn1_1.99 -NCPU=6 -TCPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-327-750D-30
Query length: 128
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 146.090000

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-381-881-1	+	93.50	169.00	0.0919	1315
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-281-221-1	+	93.50	169.00	0.0919	1315
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-931-999-4	+	86.50	137.69	5.09	6755
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-809-740A-1	-	85.00	142.20	2.86	3252
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-809-740A-4	-	85.00	142.20	2.86	3252
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-426-599B-3	+	83.50	156.25	0.4714	624
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-505-553C-6	+	82.50	154.49	0.5910	607
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-030-995-1	+	82.50	146.08	1.74	1365
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-403-852D-1	+	81.50	129.79	14.04	5392
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-510-646B-1	+	81.50	129.79	14.04	5392
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-231-818-1	+	81.00	129.79	14.04	5392
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-948-705-1	+	81.00	141.68	3.05	1552
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-393-985-17	+	81.00	138.85	4.39	2040
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-951-715A-6	+	81.00	132.89	9.44	3624
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-459-448A-6	+	81.00	132.89	9.44	3624
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-459-504B-6	+	81.00	132.89	9.44	3624
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-459-504B-6	+	81.00	132.89	9.44	3624
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-053-549-7	+	81.00	132.89	9.44	3624
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-717-294-42	+	79.50	127.68	18.40	4451
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-053-549-1	+	79.00	119.52	52.37	8854
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-147-777-3	+	78.50	107.74	237.26	24979
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-452-872-3	+	78.50	107.74	237.26	24979
/cgn2_6/ptodata/2/ina/PTUS_COMB.seq:US-09-03985-3	+	78.50	107.74	237.26	24979
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-660-645A-4	-	77.50	140.07	3.76	908
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-298-718-4	-	77.50	140.07	3.76	908
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-546-969-4	-	77.50	140.07	3.76	908
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-980-832-1	+	77.50	116.72	75.00	8625
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-980-832-27	+	77.50	113.98	106.58	11233
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-764-233A-1	+	77.00	97.60	870.99	49377
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-452-075-1	-	76.50	126.19	22.27	2841
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-231-061-1	-	76.50	126.19	22.27	2841
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-402-282-3	+	76.50	108.49	215.70	15664
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-508-004-3	+	76.50	108.49	215.70	15664
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-402-066-3	+	76.50	108.49	215.70	15664
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-402-068-3	+	76.50	108.49	215.70	15664
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-592-874-1	-	76.50	102.17	484.98	28804
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-096-942-2	-	76.50	102.17	484.98	28804
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-096-867-2	-	76.50	102.17	484.98	28804
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-258-261B-6	+	76.50	102.12	488.43	28958

/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-456-837-6 + 76.50 102.12 488.43 28958
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-457-342-6 + 76.50 102.12 488.43 28958
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-457-646A-6 + 76.50 102.12 488.43 28958
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-458-076A-6 + 76.50 102.12 488.43 28958
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-764-233A-4 + 76.50 102.12 488.43 28958

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-381-881-1

seq_documentation_block:
; Sequence 1, Application US/08381881
; Patent No. 5939064
; GENERAL INFORMATION:
; APPLICANT: SAVALKOL, PAUL H.M.
; APPLICANT: GAASTRA, WILLEM
; TITLE OF INVENTION: BORDETELLA BRONCHISEPTICA VACCINE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STEVENS, DAVIS, MILLER & MOSHER
; STREET: 515 NORTH WASHINGTON STREET.
; CITY: ALEXANDRIA
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/381,881
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: POULOS III, JAMES A.
; REGISTRATION NUMBER: 31,714
; REFERENCE/DOCKET NUMBER: TPP 29685
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703 549-7200
; TELEFAX: 703 528-5313
; TELEX: 44-0704 "STEVENS"
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1315 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bordetella bronchiseptica
; STRAIN: 401
; IMMEDIATE SOURCE:
; CLONE: E coli PC2495(pivb3-420)
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..539
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 540..1142
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1143..1315
; US-08-381-881-1

alignment_scores:
Quality: 93.50 Length: 96
Ratio: 1.908 Gaps: 4
Percent Similarity: 51.042 Percent Identity: 33.333
alignment_block:
US-09-327-750D-30 x US-08-381-881-1

```

Align seg 1/1 to: US-08-381-881-1 from: 1 to: 1315
43 AlaGlyLysAsnCysAlaProArgGlyGlyArgArgPheArgValar 59
542 GCAAGCCCAAAACGTTCTCTCGGCGCGCTCGCGCGCTCGCGCTCG 591
59 gGlnProLeu.....AlaHisTyrArgTrrp..... 67
592 CGGCCCATCGCGAAGACCGCACCATTTGCTATTACCGCGCACGAC 641
68 ..AspLeuMetGlnArgValGlyGluPro.....GlnGlyArgMet 80
642 CAGACCTGCAGATCGAGGACCGCGCGCTTACATCAAGTGTGCA 691
81 ArgGluGlnAsnValGlnArgPheGlyGlyAspValArgGlnLeuMetG1 97
692 CTGCCCCAC.....GATCTCCAAGAGCGCGCTGA 720
97 uLysLeuArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThra 114
721 AGAAGCCCGCGGACGTCGCGGCGCGCTCGCTTCGATATCAAGCTGAAG 770
114 spProProHisHisAspHisAspGluPheCysLeu 126
771 GACTGCCCGACCGCTCAACACTCTCAAGCTGTACTT 808
seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-09-281-221-1
seq_documentation_block:
; Sequence 1, Application US/09281221
; Patent No. 6284256
; GENERAL INFORMATION:
; APPLICANT: SAELEKOU, PAUL H.,M.
; TITLE OF INVENTION: BORDETTELLA BRONCHISEPTICA VACCINE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STEVENS, DAVIS, MILLER & MOSHER
; STREET: 515 NORTH WASHINGTON STREET.
; CITY: ALEXANDRIA
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/281,221
; FILING DATE: 30-Mar-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/381,881
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: POULOS III, JAMES A.
; REGISTRATION NUMBER: 31,714
; REFERENCE/DOCKET NUMBER: TPP 29685
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703 549-7200
; TELEFAX: 703 528-5313
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

```

```

ORIGINAL SOURCE:
ORGANISM: Bordetella bronchiseptica
STRAIN: 401
IMMEDIATE SOURCE:
CLONE: E coli PC2495(pIVB3-420)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..539
FEATURE:
NAME/KEY: CDS
LOCATION: 540..1142
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1143..1315
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-281-221-1

alignment_scores:
Quality: 93.50 Length: 96
Ratio: 1.908 Gaps: 4
Percent Similarity: 51.042 Percent Identity: 33.333

alignment_block:
US-09-327-750D-30 x US-09-281-221-1 ..
Align seg 1/1 to: US-09-281-221-1 from: 1 to: 1315
43 AlaGlyLysAsnCysAlaProArgGlyGlyArgArgPheArgValar 59
542 GCAAGCCCAAAACGTTCTCTCGGCGCGCTCGCGCGCTCGCGCTCG 591
59 gGlnProLeu.....AlaHisTyrArgTrrp..... 67
592 CGGCCCATCGCGAAGACCGCACCATTTGCTATTACCGCGCACGAC 641
68 ..AspLeuMetGlnArgValGlyGluPro.....GlnGlyArgMet 80
642 CAGACCTGCAGATCGAGGACCGCGCGCTTACATCAAGTGTGCA 691
81 ArgGluGlnAsnValGlnArgPheGlyGlyAspValArgGlnLeuMetG1 97
692 CTGCCCCAC.....GATCTCCAAGAGCGCGCTGA 720
97 uLysLeuArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThra 114
721 AGAAGCCCGCGGACGTCGCGGCGCGCTCGCTTCGATATCAAGCTGAAG 770
114 spProProHisHisAspHisAspGluPheCysLeu 126
771 GACTGCCCGACCGCTCAACACTCTCAAGCTGTACTT 808
seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-931-999-4
seq_documentation_block:
; Sequence 4, Application US/08931999
; Patent No. 6043219
; GENERAL INFORMATION:
; APPLICANT: Iandolo, John J.
; APPLICANT: Crupper, Scott S.
; TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hovey, Williams, Timmons & Collins
; STREET: 2405 Grand Boulevard, Suite 400
; CITY: Kansas City
; STATE: Missouri
; COUNTRY: U.S.A.
; ZIP: 64108
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```


us-09-327-750d-30.rni

Tue Mar 12 09:01:30 2002

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/426,599B
; FILING DATE: 21-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saxe, Stephen A.
; REGISTRATION NUMBER: 38,609
; REFERENCE/DOCKET NUMBER: 1604-123A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-624-1589
; TELEFAX: 202-783-6031
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 624 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Optimized cDNA"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..621
; US-08-426-599B-3

alignment_scores:
    Quality: 83.50      Length: 96
    Ratio: 1.670       Gaps: 6
    Percent Similarity: 52.083      Percent Identity: 34.375

alignment_block:
    US-09-327-750D-30 x US-08-426-599B-3 ..

Align seg 1/1 to: US-08-426-599B-3 from: 1 to: 624
33 ArgGluProAlaValAlaLeuLeuSerGluAlaGlyLysAsnCysAlaPr 49
|||||
15 CGTCAACG.....CTGCTCTACACCGCTGGCGCGCGCCCTCCAA... 56
|||||
49 oArgGlyValArgArgPheArgValArgValArgProIleAlaHisTyrA 66
|||||
57 .....GGGCGACGCGCGCTCGACGCGCGCGCGCGCGCACTCACTCC 99
|||||
66 rGTrpAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGlu 82
|||||
100 GG.....CGAGTCTGGACCATCAACGTCGAGCCCGGCAC 134
|||||
83 GluAsnValGlnArgPheGlyValArgValArgGlnLeuMetGluLysLe 99
|||||
135 CAAGCGCGCAAGATCTGGCGCGCAC...CGACTGCTACTTCGACGACT 181
|||||
99 uArgGluArgGlnLeuSerHis..... 106
|||||
182 CCGCGCGCGCAATCTCGCGCACCGCGACTGCGCGCGCTCTCTCAAGTGC 231
|||||
107 ..SerLeuArgAlaValSerThrAspProProHisHis 118
|||||
232 AAGCGCTCG.....CGCGCGCGCGCACAC 257
|||||

seq_name: /cqn2.6/ptodata/2/1na/6A_COMB.seq:US-08-506-553C-6

seq_documentation_block:
; Sequence 6, Application US/08506553C
; Patent No. 6120989
; GENERAL INFORMATION:
; APPLICANT: Vornhagen, Rolf; Hinderer, Walter; Sonneborn, Han-H.;
; APPLICANT: Plachter, Bodo; and Jahn, Gerhard
; TITLE OF INVENTION: ISOLATED HUMAN CYTOMEGALOVIRUS
; POLYPEPTIDES AND USES THEREOF

```

```

; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/506,553C
; FILING DATE: 07/25/95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary Anne Schofield
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LEDEP 202-PFF/MAS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 607 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-08-506-553C-6

alignment_scores:
    Quality: 82.50      Length: 117
    Ratio: 1.650       Gaps: 4
    Percent Similarity: 42.735      Percent Identity: 26.496

alignment_block:
    US-09-327-750D-30 x US-08-506-553C-6 ..

Align seg 1/1 to: US-08-506-553C-6 from: 1 to: 607
50 ArgGlyValArgArgPheArgValArgGlnProIleAlaHisTyrA 66
|||||
107 CGGCTCGGAGGAGGTGGCGCGGTGGCGCGCAAGAA...GCACGACCGG 153
|||||
66 gTrpAspLeuMetGlnArgValGlyGluProGln..... 77
|||||
154 GTGGCGCGGTGGTTCGGTACGCGGAAATGAGTAGCGGTGGCGCGGC 203
|||||
77 ..... 77
|||||
204 GGTATCATGACACCGTCTTTCTCCAAGGAAATACGACGACACAA 253
|||||
78 .....GlyArgMetArgGluG1 83
|||||
254 GATCACCAGCTACCTGACGTCCAAAGGTGGATCGGCGCGCGGAGGAG 303
|||||
83 uAsnValGlnArgPheGlyValArgGlnLeuMetGluLysLeuA 100
|||||
304 GAGGAGCGCGGTGGATCGCACTCGGCAATCTTCAACGACCGG 353
|||||
100 rG.....GluArgGlnLeuSer..... 105
|||||
354 AAAGAGAGAGCGACGACGAGGAGTCTGTAACTTCGAGTTCGCCCTAA 403
|||||
106 HisSerLeuArgAlaValSerThrAspProProHisHisAspHis 122
|||||

```



```

33 g.....GluProAla.....ValAlaL 39
4850 GCGGTTCCGGATCAGTATCCCGCAGCGTGGCTCATGTCTCGCTC 4899
39 euileSerGluAlaGlyLysAsnCysAlaPro.....ArgGlyGlyArg 53
4900 TCTTCTCGTCCGACGACTCAGCTGCGCGCGGGGGAGTGGCGGACAA 4949
54 ArgArgPheArgValArgGlnProIleAlaHisTyrArgTrpAspLeu 70
4950 CGTCGCTTCGCGCCCTCGGCGCTCAGGACACCGCGCTCGAC 5057
70 tGlnArgValGlyGluProGlnGlyArgMetArgGluGluAsnValGln 87
4983 .....CCCCAGCTGGTGGCGCGCTGGCGGACGCGCTCGCC 5019
87 rgPheGlyGlyAspValArgGlnLeuMetGluLysLeuArgGluArgGln 103
5020 GA.....GTACGAGGTCTGGCTCAAGAACACCGCGCTCGACAC 5057
104 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspH 120
5058 CGGCCCCGCTCGTGGTCTCCACCGAGCGGACACCGCGCTCATGTGCA 5107
120 sHis 121
5108 TCAC 5111
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-510-646B-1
seq_documentation_block:
; Sequence 1, Application US/08510646B
; Patent No. 6077699
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; APPLICANT: Blanche, Francis
; APPLICANT: Crouzet, Joel
; APPLICANT: Jacques, Nathalie
; APPLICANT: Lacroix, Patricia
; APPLICANT: Thibaut, Denis
; APPLICANT: Zagorec, Monique
; APPLICANT: Debusche, Laurent
; APPLICANT: De Crecy-Lagard, Valerie
; TITLE OF INVENTION: Polypeptides Involved In The
; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
; TITLE OF INVENTION: Coding For These Polypeptides And Their Use
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fimegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/510,646B
; FILING DATE: 03-AUG-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,852
; FILING DATE: 10-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR 93/00923
; FILING DATE: 25-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/11441
; FILING DATE: 25-SEP-1992
; ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5392 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sp. pristinacspiralis
US-08-510-646B-1
alignment_scores:
Quality: 81.50 Length: 118
Ratio: 1.430 Gaps: 5
Percent Similarity: 48.305 Percent Identity: 27.966
alignment_block:
US-09-327-750D-30 x US-08-510-646B-1
Align seg 1/1 to: US-08-510-646B-1 from: 1 to: 5392
17 HisGlnLysLysGluGluLysGluLysProGlnAspThrIleArg 33
4800 CATCGTGTCCACCGTTCATCCGACCGGACCATCTGATGGTCTCCCG 4849
33 g.....GluProAla.....ValAlaL 39
4850 GCGGTTCCGGATCAGTATCCCGCAGCGTGGCTCATGTCTCGCTC 4899
39 euileSerGluAlaGlyLysAsnCysAlaPro.....ArgGlyGlyArg 53
4900 TCTTCTCGTCCGACGACTCAGCTGCGCGGGGGAGTGGCGGACAA 4949
54 ArgArgPheArgValArgGlnProIleAlaHisTyrArgTrpAspLeu 70
4950 CGTCGCTTCGCGCCCTCGGCGCTCAGGACACCGCGCTCGAC 5057
70 tGlnArgValGlyGluProGlnGlyArgMetArgGluGluAsnValGln 87
4983 .....CCCCAGCTGGTGGCGCGCTGGCGGACGCGCTCGCC 5019
87 rgPheGlyGlyAspValArgGlnLeuMetGluLysLeuArgGluArgGln 103
5020 GA.....GTACGAGGTCTGGCTCAAGAACACCGCGCTCGACAC 5057
104 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspH 120
5058 CGGCCCCGCTCGTGGTCTCCACCGAGCGGACACCGCGCTCATGTGCA 5107
120 sHis 121
5108 TCAC 5111
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-231-818-1
seq_documentation_block:
; Sequence 1, Application US/09231818
; Patent No. 6171846
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; APPLICANT: Blanche, Francis
; APPLICANT: Crouzet, Joel
; APPLICANT: Jacques, Nathalie
; APPLICANT: Lacroix, Patricia

```


[illegible]

```

seq_documentation_block:
; Sequence 6, Application US/08459448A
; Patent No. 5859336
;
; GENERAL INFORMATION:
;
; APPLICANT: Koziel, Michael G.
; APPLICANT: Desai, Nallini M.
; APPLICANT: Lewis, Kelly S.
; APPLICANT: Kramer, Vance C.
; APPLICANT: Warren, Gregory W.
; APPLICANT: Evola, Stephen V.
; APPLICANT: Crosland, Lyle D.
; APPLICANT: Wright, Martha S.
; APPLICANT: Merlin, Ellis J.
; APPLICANT: Maunis, Karen L.
; APPLICANT: Rothstein, Steven J.
; APPLICANT: Bowman, Cindy G.
; APPLICANT: Dawson, John L.
; APPLICANT: Dunder, Erik M.
; APPLICANT: Pace, Gary M.
; APPLICANT: Suttie, Janet L.
;
; TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
;
; TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
;
; NUMBER OF SEQUENCES: 94
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: No. 3859336artis Corporation
; STREET: Patent & Trademark Dept., 520 White Plains
; STREET: Rd., POB 2005

```

DATE: 10/24/2003
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/459,448A
FILING DATE: 02-JUN-1995

FILING DATE: 02-JUN-1995
CLASSIFICATION: 800

FILING DATE: 25-SEP-1992
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/772,027
FILING DATE: 04-OCT-1991

ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.

REGISTRATION NUMBER: 40403
REFERENCE/DOCKET NUMBER: CGC 1577/

REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (818) 541-9592

TELEPHONE: (919) 541-8582
TELEFAX: (919) 541-8689

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:

LENGTH: 3624 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE:	other nucleic acid
DESCRIPTION:	/desc = "Synthetic DN
HYDROTHERMAL:	NO

HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS

NAME/KEY: CDS
LOCATION: 1..3621
OTHER INFORMATION: /product= "Full

OTHER INFORMATION: /product= "full"
OTHER INFORMATION: optimized cryib"
OTHER INFORMATION: /note= "disclos

OTHER INFORMATION;
-459-448A-6

alignment_scores:

OM of: US-09-327-750D-31 to: EST.* out_format : pfs
 Date: Mar 11, 2002 2:16 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
 -MODEL=framet_p2n.model -DEV=xlh
 -Q/cgnt2/_USPTO_spoof/US09327750/runat_11032002_101153_20308/app_query.fasta.1.1472
 -DB=EST -QFWT=fastap -SUFFIX=1st -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOAPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZ=500
 -MINLEN=0 -MAXLEN=200000000 -USER=US09327750 -ECGN1_1.5654
 -NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
 Query: US-09-327-750D-31
 Query length: 128
 Database: EST.*
 Database sequences: 11351937
 Database length: 107921985
 Search time (sec): 4085.940000

score_list:	Strd Orig	zScore	EScore	Len	Documentation
gb_est1:BE314909	+	679.00	1289.59	1.1e-62	594
gb_est1:AL520931	-	679.00	1287.82	1.4e-62	717
gb_est1:BE889313	+	679.00	1287.12	1.6e-62	717
gb_est1:BG709503	+	679.00	1286.67	1.7e-62	810
gb_est1:BG707398	+	679.00	1286.38	1.7e-62	835
gb_est1:BF038787	+	679.00	1286.29	2.3e-62	1042
gb_est1:BG709503	+	679.00	1286.25	2.3e-62	835
gb_est1:BG707398	+	679.00	1286.25	2.3e-62	835
gb_est1:BG714974	+	672.00	1274.33	8.1e-62	726
gb_est1:BF967675	+	672.00	1273.49	9.0e-62	793
gb_est1:BG707398	+	672.00	1273.21	9.4e-62	817
gb_est1:BG707398	+	668.00	1265.62	2.5e-61	813
gb_est1:AI291270	+	668.00	1266.17	2.3e-61	511
gb_est1:AI291126	+	664.00	1259.81	5.2e-61	669
gb_est1:BF237433	+	664.00	1259.24	5.6e-61	711
gb_est1:BE366012	+	663.00	1258.77	6.0e-61	610
gb_est1:BE33217	+	663.00	1251.32	1.6e-60	897
gb_est1:BF967457	+	658.00	1247.04	2.7e-60	769
gb_est1:BG715745	+	658.00	1246.48	2.9e-60	816
gb_est1:BF685596	+	657.00	1244.55	3.7e-60	818
gb_est1:BG709294	+	646.00	1223.96	5.2e-59	782
gb_est1:AL520932	+	644.00	1219.53	9.2e-59	834
gb_est1:BG709294	+	638.00	1209.80	3.2e-58	694
gb_est1:BG715353	+	636.00	1204.56	6.2e-58	807
gb_est1:BG715353	+	606.50	1151.87	5.4e-55	547
gb_est1:BG655001	+	606.50	1150.48	6.4e-55	634
gb_est1:BG712167	+	606.50	1150.13	6.7e-55	658
gb_est1:BG712167	+	606.50	1149.91	6.9e-55	674
gb_est1:BG714823	+	606.50	1149.68	7.1e-55	690
gb_est1:AV702285	+	606.50	1149.67	7.1e-55	691
gb_est1:AV705808	+	606.50	1149.62	7.2e-55	695
gb_est1:AL526209	+	606.50	1149.56	7.2e-55	699
gb_est1:BG820179	+	606.50	1148.49	8.3e-55	783
gb_est1:BE790774	+	606.50	1148.36	8.4e-55	794
gb_est1:AL526247	+	606.50	1148.01	1.5e-54	683
gb_est1:AV702643	+	599.50	1138.01	3.2e-54	576
gb_est1:AV729703	+	599.50	1135.73	4.3e-54	734
gb_est1:BG715659	+	599.50	1135.10	4.6e-54	386
gb_est1:AV732454	+	596.00	1135.82	1.3e-53	929
gb_est1:BF316327	+	596.00	1123.54	2.0e-53	585
gb_est1:AI141583	+	592.00	1117.40	4.5e-53	828
gb_est1:BG709325	+	590.50	1117.40	4.5e-53	828

gb_est1:AI826617 - 590.00 1119.74 3.3e-53 584 | AI826617 WK3503.x1 NCI_CGAF
 gb_est1:AW160832 + 584.50 1110.13 1.1e-52 531 | AW160832 au76b12.y1 Schneide
 gb_est1:BF346478 + 583.50 1106.22 1.9e-52 656 | BF346478 602020342F1 NCI_CG
 gb_est1:AV727777 + 580.50 1101.83 3.3e-52 569 | AV727777 AV727777 HTC Homo s
 gb_est1:AW953922 + 580.50 1101.23 3.6e-52 607 | AW953922 EST365887 MAGE rese
 seq_name: gb_est1:BE314909
 seq_documentation_block:
 LOCUS BE314909 594 bp mRNA EST 26-OCT-2000
 DEFINITION 601140074F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3049290 5',
 mRNA sequence.
 ACCESSION BE314909
 VERSION BE314909.1 GI:9145006
 KEYWORDS EST.
 SOURCE human;
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 594)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaaps-r@mail.nih.gov
 Tissue Procurement: DCTD/DTP
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: L1CM93 row: c column: 19
 High quality sequence stop: 594.
 Location/Qualifiers
 1..594
 ..organism="Homo sapiens"
 ..db_xref="taxon:9606"
 ..clone="IMAGE:3049290"
 ..clone_lib="NIH_MGC_9"
 ..tissue_type="adenocarcinoma cell line"
 ..lab_host="DH10B (phage-resistant)"
 ..note="Organ: ovary; Vector: pORF7; Site:1; XhoI; Site:2;
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GCCACGAG(G). Library constructed by Ling Hong in
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 161 a 126 c 181 g 126 t
 ORIGIN
 alignment_scores:
 Quality: 679.00 Length: 128
 Ratio: 5.305 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-327-750D-31 x BE314909
 Align seg 1/1 to: BE314909 from: 1 to: 594
 1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuLeuValGluAsnVa 17
 141 ATGGAGTCCAAAGACGAGCGGTTAAACATCTCATCGTGGAAATGT 190
 17 LasnGluGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
 191 CAACAGGAAATGATGATGAAAGATGAAAGGACGAGTTCGCTATAAAG 240
 34 LyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50

Align seg 1/1 to reverse of: AL520931 from: 1 to: 717

```
1 MetGluSerIysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
530 ATGGAGTGCCAAAGAGGAACAGCGTTAAACAATCTCATCGTGGAAAAATGT 481
|||||
17 IasnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
480 CAACAGCGAAAAATGATGAAAAAGATGAAAAGAGCAGAGTGTCTAATAAAG 431
|||||
34 LYGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
430 GGGAGCCCTTGGCCCTACCTTTGAATGTTAGTGAATACTGTGTGCCTAGA 381
|||||
51 GlyAsnArgArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
380 GGAACACCGTAGGCGGGTTCGCGGTTAGGCAGGCCCATCTTCGCAGTATAGATG 331
|||||
67 pAspIleMethHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84
330 GGACATAATGCATAGGCGTTGGAGAGCCACAGCAAGGATGAGAGAGGAGA 281
|||||
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
|||||
280 ATATGGAAAGGATGCGGGAGGAGGTGAGCACACTGATGGAAAACGCTGAGG 231
|||||
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH1 117
230 GAAAAGCAGTTGATCATATCTCGCGGCAGTCAGCACTGATGCCCTCA 181
|||||
117 shisAspHisHisAspGluPheCysLeuMetPro 128
180 CCATGACCATCACGATGAGTTTGGCTTATGSCC 147
|||||
```

17c-60-60

```

seq_name: gb_est2:BE889313

sseq_documentation_block:
  BLOCUS BE889313 772 bp mRNA EST 20-OCT-2000
  DEFINITION 601513201F1 NTH_MGC_71 Homo sapiens cDNA clone IMAGE:3914401 5',
  mRNA sequence.
  BE889313
  BE889313 BE889313.1 GI:1034503
  EST.
  SOURCE
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 772)
  NIH-MGC http://mgi.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: c9apbs-re@mail.nih.gov
  Tissue Procurement: ATCC
  CDNA Library Preparation: Life Technologies, Inc.
  CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Incyte Genomics, Inc.
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LLN9736 row: b column: 02
  High quality sequence stop: 763.
  Location/Qualifiers
    1..772
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:3914401"
      /clone_lib="NTH_MGC_71"
      /tissue_type="leptomysarcoma"
      /lab_host="DH10B (phage-resistant)"
  Note: 1: Organ: uterus; Vector: pCMV-SPORT6; Site: 1: NotI;
  Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT

```

AA1/TCE07CTW X TC-DNC / -17C-60-60

http://image.llnl.gov
Plate: LLAM10683 row: a column: 22
High quality sequence stop: 770.
Location/Qualifiers
1. 810
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4797213"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/notes="Organ: Brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcggg); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 206 a 168 c 247 g 187 t 2 others
ORIGIN
alignment_scores
Quality: 679.00 Length: 128
Ratio: 5.305 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-327-750D-31 x BG709503
Align seg 1/1 to: BG709503 from: 1 to: 810
1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
218 ATGGAGTCCAAAGAGGAGGAGCGTTAAACAATCTCATCTGGGAAATGT 267
17 lAsnGlnGluAsnAspGluLysAspGluGlnValAlaAsnLysG 34
268 CAACCAAGAAATGATGAAAGATGAGAGGAGCAAGTCTCTAATAAG 317
34 lYgluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
318 GGGAGCCCTTGGCCCTACCTTTGAATGTAGTGAATCTGTGCTTGA 367
51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
368 GGAACCGTAGCGGTTCCCGTTAGGAGCCCATCTCGAGTATAGATG 417
67 pAspTleuMetHisArgLeuGluGluProGlnAlaArgMetArgGluGluA 84
418 GGACATATATCATAGGCTTGGAGAGCCACAGGAGGATGAGAGGAGA 467
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
468 ATATGGAAGGATTTGGGAGGAGTGCAGACGCTGATGGAAGCTGAGG 517
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHI 117
518 GAAAGCAGTTGAGTCTATAGTCTGCGGCGAGTGCAGACTGATCCCTCA 567
117 sHisAspHisAspGluPheCysLeuMetPro 128
568 CCATGACCATCAGATGAGTTTGCCTTATGCC 601
seq_name: gb_est2:BG707398
seq_documentation_block:
LOCUS BG707398 835 bp mRNA
DEFINITION 602672810F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4795347 5',
mRNA sequence.
ACCESSION BG707398
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cspbs-r@mail.nih.gov
Tissue Procurement: Miklos Falkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

us-09-327-750d-31 x BE889313
Average insert size 2.1 kb.
226 a 139 c 222 g 185 t
BASE COUNT
ORIGIN
alignment_scores
Quality: 679.00 Length: 128
Ratio: 5.305 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-327-750D-31 x BE889313
Align seg 1/1 to: BE889313 from: 1 to: 772
1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
125 ATGGAGTCCAAAGAGGAGGAGCGTTAAACAATCTCATCTGGGAAATGT 174
17 lAsnGlnGluAsnAspGluLysAspGluGlnValAlaAsnLysG 34
175 CAACCAAGAAATGATGAAAGATGAAAGGAGCAAGTGTCTAATAAG 224
34 lYgluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
225 GGGAGCCCTTGGCCCTACCTTTGAATGTAGTGAATCTGTGCTTGA 274
51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
275 GGAACCGTAGCGGTTCCCGTTAGGAGCCCATCTCGAGTATAGATG 324
67 pAspTleuMetHisArgLeuGluGluProGlnAlaArgMetArgGluGluA 84
325 GGACATATATCATAGGCTTGGAGAGCCACAGGAGGATGAGAGGAGA 374
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
375 ATATGGAAGGATTTGGGAGGAGTGCAGACAGCTGATGGAAGCTGAGG 424
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHI 117
425 GAAAGCAGTTGAGTCTATAGTCTGCGGCGAGTGCAGACTGATCCCTCA 474
117 sHisAspHisAspGluPheCysLeuMetPro 128
475 CCATGACCATCAGATGAGTTTGCCTTATGCC 508
seq_name: gb_est2:BG709503
seq_documentation_block:
LOCUS BG709503 810 bp mRNA
DEFINITION 602674781F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4797213 5',
mRNA sequence.
ACCESSION BG709503
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cspbs-r@mail.nih.gov
Tissue Procurement: Miklos Falkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

```

VERSION      BG707398.1  GI:13983707
KEYWORDS     EST.
SOURCE       Human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 835)
AUTHORS      NIH-MGC http://mgi.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgaaps-r@mail.nih.gov
              Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
              cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
              Toshlyuki and Piero Carninci (RIKEN)
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LHM10678 row: d column: 04
              High quality sequence stop: 814.
              Location/Qualifiers
              1..835
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone_image="4795347"
              /clone_lib="NIH_MGC_96"
              /tissue_type="hypothalamus"
              /lab_host="DH108"
              /note="Organ: brain; Vector: pBluescriptR (modified
              pluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcag
              ); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3',
              size-selected for average insert size 2.3 kb and
              normalized to 50x. This is a primary library enriched
              for full-length clones and constructed using the
              Cap-trapper method (Carninci, in preparation). Library
              constructed by M. Brownstein (NHGRI), National
              Institutes of Health). Note: this is a NIH_MGC Library."
              220 a 176 c 246 g 193 t
              BASE COUNT
              ORIGIN

alignment_scores:
  Quality: 679.00      Length: 128
  Ratio: 5.305        Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-327-750d-31 x BG707398 ...
Align seg 1/1 to: BG707398 from: 1 to: 835
1 MetGluSerLysGluGluArgAlaLeuAsnLeuValGluAsnVa 17
|||||
248 ATGGAGTCCAAAGAGGACGAGCGTTAAACAATCTCATCTGTTGAAAATGT 297
|||||
17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
|||||
298 CAACGAGCAAAATGATGAAAGATGAAAGGAGCAAGTTGCTATTAAG 347
|||||
34 lYGlupProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
348 GGGAGCCCTTGGCCCTACCTTGAATGTTAGTGAATACTGTGTGCCTAGA 397
|||||
51 GlyAsnArgArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
|||||
398 GGAACACCTAGGCGTTCCGGTTAGGACGCCCATCTCTGCAGTATAGTG 447
|||||
67 pAspIleMetHisArgLeuGluGluProGlnAlaArgMetArgGluGlu 84
|||||
448 GGACATAATGATAGGTTGGAGAGCCACAGGCAAGGATGATGAGAGGAGA 497
|||||

```

```

84 smMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
|||||
498 ATATGCAAGAGGATTGGGGAGGAGGTGACACAGCTGATGGAAAGCTGAGG 547
|||||
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 117
|||||
548 GAAAGCACTTGGTTCATAGTCTGCGGCGAGTCAGCAGCTGATCCCTCA 597
|||||
117 sHisAspHisHisAspGluPheCysLeuMetPro 128
|||||
598 CCATGACCATCACGATGAGTTTTCCTTATGCC 631
|||||
seq_name: gb_est2:BF038787
seq_documentation_block:
LOCUS      BF038787      1042 bp      mRNA      20-OCT-2000
DEFINITION 601462119F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3865348 5',
            mRNA sequence.
ACCESSION  BF038787
VERSION    BF038787.1 GI:10745987
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1042)
AUTHORS    NIH-MGC http://mgi.nci.nih.gov/
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaaps-r@mail.nih.gov
            Tissue Procurement: DCTD/DTF
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LHM9608 row: f column: 05
            High quality sequence stop: 732.
            Location/Qualifiers
            1..1042
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_image="3865348"
            /clone_lib="NIH_MGC_66"
            /tissue_type="adenocarcinoma"
            /lab_host="DH108 (phage-resistant)"
            /note="Organ: ovary; Vector: pCMV-SpORT6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
            Average insert size 1.8 kb. Library constructed by Life
            Technologies."
            308 a 200 c 296 g 237 t 1 others
            BASE COUNT
            ORIGIN

alignment_scores:
  Quality: 679.00      Length: 128
  Ratio: 5.305        Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-327-750d-31 x BF038787 ..
Align seg 1/1 to: BF038787 from: 1 to: 1042
1 MetGluSerLysGluGluArgAlaLeuAsnLeuValGluAsnVa 17
|||||
85 ATGGAGTCCAAAGAGGACGAGCGTTAAACAATCTCATCTGTTGAAAATGT 134
|||||
17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
|||||
135 CAACGAGCAAAATGATGAAAGATGAAAGGAGCAAGTTGCTATTAAG 184
|||||

```

34 lyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
185 GGGAGCCCTGGCCCTACTTGAATGTTAGTGAATCTACTGTGCTCCCTAGA 234
51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
235 GGAACCGTAGCGGTTCGGGTTAGGCAGCCCATCTCGCAGTATAGATG 284
67 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84
285 GGACATAATGATAGCTTGGAGAGCCACAGGCAAGGATGAGAGGAGA 334
84 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 100
335 ATATGAAAGGATCGGGGAGGAGGTGACACAGCTGATGGAAGCTGAGG 384
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 117
385 GAAAGCAGTTGAGTCATAGTTTGGGGCAGTCAGCAGCTGATCCCTCA 434
117 sHisAspHisHisAspGluPheCysLeuMetPro 128
435 CCATGACCATCAGGATGAGTTTGGCTTATGCC 468
seq_name: gb_est2:BG285825

seq_documentation_block:
LOCUS BG285825 946 bp mRNA 21-FEB-2001 EST
DEFINITION 602380917F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:449845 5',
mRNA sequence.
ACCESSION BG285825
VERSION BG285825.1 GI:13038171
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 946)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10361 row: a column: 06
High quality sequence stop: 710.
Location/Qualifiers
1..946
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:449845"
/clone_lib="NIH_MGC_93"
/tissue_type="transitional cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: bladder; Vector: pCMV-Sport6; Site: 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 276 a 201 c 279 g 190 t
ORIGIN
alignment_scores:
Quality: 674.00 Length: 128
Ratio: 5.266 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 99.219
alignment_block:
US-09-327-750D-31.x BG285825
Align seg 1/1 to: BG285825 from: 1 to: 946
1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
190 ATGAGTCCAAAGAGGAGGAGCGGTTAAACAATCTCATCGTGGAAAAATGT 239
17 lAsnGlnGluAsnAspGluLysAspGluLysGlnValAlaAsnLysG 34
240 CAACGAGGAAAAATGATGAAAAAGATGAAAAAGGCAAGTTGCTAATAAG 289
34 lyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
290 GGGAGCCCTGGCCCTACTTTCATGTTAGTGAATCTACTGTGCTCCCTAGA 339
51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
340 GGAACCGTAGCGGTTCGGGTTAGGCAGCCCATCTCGCAGTATAGATG 389
67 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84
390 GGACATAATGATAGCTTGGAGAGCCACAGGCAAGGATGAGAGGAGA 439
84 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 100
440 ATATGGAAGGATCGGGGAGGAGGTGAGCAGCTGATGGAAGCTGAGG 489
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 117
490 GAAAGCAGTTGAGTCATAGTTTGGGGCAGTCAGCAGCTGATCCCTCA 539
117 sHisAspHisHisAspGluPheCysLeuMetPro 128
540 CCATGACCATCAGGATGAGTTTGGCTTATGCC 573
seq_name: gb_est2:BG714974

seq_documentation_block:
LOCUS BG714974 726 bp mRNA 08-MAY-2001 EST
DEFINITION 602677263F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4799823 5',
mRNA sequence.
ACCESSION BG714974
VERSION BG714974.1 GI:13993905
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 726)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10689 row: n column: 16
High quality sequence stop: 723.
Location/Qualifiers
1..726
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4799823"

FEATURES
source
1..726

```

/clone_lib="NIH_MGC_96"
/tissue_type="hypothalmus"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.3 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT      181 a      159 c      226 g      160 t
ORIGIN

alignment_scores:
  Quality: 672.00      Length: 127
  Ratio: 5.291      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-327-750D-31 x BG714974

Align seg 1/1 to: BG714974 from: 1 to: 726

1 MetGluSerLySGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
|||||
219 ATGAGTCCAAAGAGAGAGAGCGTTAAACAATCTCATCGTGGAAAATGT 268
|||||
17 lAsnGlnGluAsnAspGluLySGluGluValAlaAlaAsnLySG 34
|||||
269 CAACAGGAAATGATGAAAAGATGAAAAGAGCAAGTTCCTAATAAAG 318
|||||
34 lGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
319 GGGAGCCCTTGCCCTTACCTTTGAATGTAGTGAATCTGTCCTAGTA 368
|||||
51 GlyAsnArgArgPheArgValArgGlnProLeuGlnTyrArgTr 67
|||||
369 GGAACCGTAGGCGGTTCGCGCTTAGCGAGCCCATCTCGCAGTATAGT 418
|||||
67 pAspIleMetHisArgLeuGluGluProGlnAlaArgMetArgGluGlu 84
|||||
419 GGCATATGTCATAGCTTGGAGACCCACAGCAGGAGGATGAGAGGAGA 468
|||||
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLySGluArg 100
|||||
469 ATATGAAAGGATTTGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAG 518
|||||
101 GluLySGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
|||||
519 GAAACAGTGTGATCATAGTCTCGGCGCAGTCAGCACTGATCCCTCA 568
|||||
117 sHisAspHisAspGluPheCysLeuMet 127
|||||
569 CCATGACCATCAGTATGAGTTTGCCCTTATG 599
|||||

seq_name: gb_est2:BF967675

seq documentation_block:
LOCUS      BF967675      793 bp      mRNA      EST      23-JAN-2001
DEFINITION  60228738371 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4374549 3',
            mRNA sequence.
ACCESSION  BF967675
VERSION    BF967675.2 GI:12388141
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 793)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/

```

TITLE JOURNAL COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
On Jan 16, 2001 this sequence version replaced gi:12334890.
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLAM10038 row: f column: 22
High quality sequence start: 20
High quality sequence stop: 762.

FEATURES Source

1..793
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4374549"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalmus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.3 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 185 a 235 c 170 g 203 t
ORIGIN

alignment_scores:
 Quality: 672.00 Length: 127
 Ratio: 5.291 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-327-750D-31 x BF967675/rev

Align seg 1/1 to reverse of: BF967675 from: 1 to: 793

1 MetGluSerLySGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
|||||
577 ATGAGTCCAAAGAGAGAGCGTTAAACAATCTCATCGTGGAAAATGT 528
|||||
17 lAsnGlnGluAspGluLySGluGluValAlaAlaAsnLySG 34
|||||
527 CAACAGGAAATGATGAAAAGATGAAAAGAGCAAGTTCCTAATAAAG 478
|||||
34 lGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
477 GGGAGCCCTTGCCCTTACCTTTGAATGTAGTGAATCTGTCCTAGA 428
|||||
51 GlyAsnArgArgPheArgValArgGlnProLeuGlnTyrArgTr 67
|||||
427 GGAACCGTAGGCGGTTCGCGCTTAGCGAGCCCATCTCGCAGTATAGATG 378
|||||
67 pAspIleMetHisArgLeuGluGluValArgGlnAlaArgMetArgGluGlu 84
|||||
377 GGACATAATGCATAGGCTTGGAGAGCCACAGCAAGGATGAGAGGAGA 328
|||||
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLySGluArg 100
|||||
327 ATATGAAAGGATTTGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGG 278
|||||
101 GluLySGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
|||||

277 GAAAGCACTTGCATAGTCTCGGGCAGTCAGCACTGATCCCTCA 228
117 shIsAspHisHisAspGluPheCysLeuMet 127
|||||
227 CCATGACCATCAGCATGAGTTTCCCTTATG 197
seq_name: gb_est2:BG707734
seq_documentation_block: 817 bp mRNA EST 07-MAY-2001
LOCUS BG707734
DEFINITION 602671235F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4794061 5',
mRNA sequence.
ACCESSION BG707734
VERSION BG707734.1 GI:13984377
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 817)
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10674 row: n column: 14
High quality sequence stop: 803.
Location/Qualifiers
1. 817
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4794061"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3',
size-selected for average insert size 2.3 kb and
normalized to Rot 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 213 a 171 c 241 g 191 t
ORIGIN
1. 817
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4794061"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3',
size-selected for average insert size 2.3 kb and
normalized to Rot 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
alignment_scores:
Quality: 672.00 Length: 128
Ratio: 5.250 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.219
alignment_block:
US-09-327-750D-31 x BG707734 ..
Align seg 1/1 to: BG707734 from: 1 to: 817
1 MetGluSerLySGluGluAlaLeuAsnValSerGluTyrCysValProArg 17
|||||
221 ATGAGTCCAAAGAGAGACGCGTAAACAATCTCATCGTGGGAAATGT 270
17 lAsnGluGluAsnAspGluLySGluGluValAlaAsnLySG 34
|||||
271 CAACACGAGAAATGATGAAAAAGATGAAAGGAGCAAGTTCCTAATAAG 320

34 lYGLuProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
321 GGGAGCCCTTGGCCCTACCTTGAATGTAGTACTGTGTGCTTGA 370
51 GlyAsnArgArgPheArgValArgGlnProLeuLeuGlnTyrArgTr 67
|||||
371 GGAACCGTAGGCGGTTCGCGCTTAGGAGCCCATCTCGCAGTAGATG 420
67 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84
|||||
421 GGACATTAATGCATAGCTTGGAGACCCACAGGAGGATGAGAGAGAGA 470
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLySGluArg 100
|||||
471 ATATGGAAGGATTTGGGAGAGAGTGCAGACAGCTGATGGAAAGCTGAGG 520
101 GluLySGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 117
|||||
521 GAAAGCAGTTGAGTCATAGTCTGGGCGCAGTCAGCACTGATCCCTCA 570
117 shIsAspHisHisAspGluPheCysLeuMetPro 128
|||||
571 CCATGACAATCAGCATGAGTTTGGCTTATGCC 604
seq_name: gb_est2:BG705843
seq_documentation_block: 813 bp mRNA EST 07-MAY-2001
LOCUS BG705843
DEFINITION 602669329F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4792038 5',
mRNA sequence.
ACCESSION BG705843
VERSION BG705843.1 GI:13980593
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 813)
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10669 row: j column: 07
High quality sequence stop: 772.
Location/Qualifiers
1. 813
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4792038"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3',
size-selected for average insert size 2.3 kb and
normalized to Rot 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 207 a 168 c 243 g 194 t
ORIGIN

alignment_scores: Quality: 668.00 Length: 128
Ratio: 5.260 Gaps: 0
Percent Similarity: 99.219 Percent Identity: 99.219

alignment_block:
US-09-327-750D-31 x BG705843 ..
Align seg 1/1 to: BG705843 from: 1 to: 813:

```

1 MetGluSerLysGluGluAlaLeuAsnLeuLeuValGluAsnVa 17
|||||
218 ATGGAGTCCAAAGAGGAGCGGTTAAACAATCTCATGTGGAATGT 267
|||||
17 lAsnGlnGluAsnAspGluLysAspGluGlnValAlaAsnLysG 34
|||||
268 CAACAGGAAATGATGAAAAAGATGAAAGGAGCAAGTTGCTAATAAG 317
|||||
34 lGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
318 GAGAGCCCTTGCCCTTGAATGTTAGTGAATCTGTGCTGCTAGA 367
|||||
51 GlyAsnArgArgArgPheArgValArgGlnProLeuLeuGlnTyrArgTr 67
|||||
368 GGAACCCGTAGCGGTTCCCGTTAGGAGCCCATCTGCGAGTATAGTG 417
|||||
67 pAspIleMetHisArgLeuGluGluProGlnAlaArgMetArgGluGlu 84
|||||
418 GGACATAATGATAGCTTGGAGAGCCAGGCAAGGATGAGAGGAGAGA 467
|||||
84 sMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
|||||
468 ATATGGAAGGATGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGG 517
|||||
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProH1 117
|||||
518 GAAMGAGCTGAGTCATAGTCTCGGGCAGTCAGCACTGATCCCCCTCA 567
|||||
117 sHisAspHisAspGluPheCysLeuMetPro 128
|||||
568 CCATGACCATCAGTATGTTGCTTATGCC 601
|||||
seq_name: gb_est1:A1291270

```

seq_documentation_block:
LOCUS A1291270 511 bp mRNA EST 29-JAN-1999
DEFINITION qm16e06.x1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1882018 3', similar to SW:HG74_HUMAN Q00994 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74. ; mRNA sequence.
ACCESSION A1291270
VERSION A1291270.1 GI:3934044
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 511)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-f@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 742 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 475.
Location/Qualifiers
1..511
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1882018"
/clone_lib="NCI_CGAP_Lu5"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 152 a 103 c 155 g 100 t 1 others
ORIGIN

alignment_scores:
Quality: 666.00 Length: 128
Ratio: 5.286 Gaps: 0
Percent Similarity: 98.438 Percent Identity: 98.438

alignment_block:
US-09-327-750D-31 x A1291270 ..
Align seg 1/1 to: A1291270 from: 1 to: 511

```

1 MetGluSerLysGluGluAlaLeuAsnLeuLeuValGluAsnVa 17
|||||
83 ATGGAGTCCAAAGAGGAGCGGTTAAACAATCTCATGTGGAATGT 132
|||||
17 lAsnGlnGluAsnAspGluLysAspGluGlnValAlaAsnLysG 34
|||||
133 CAACAGGAAATGATGAAAAAGATGAAAGGAGCAAGTTGCTAATAAG 182
|||||
34 lGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
183 GGGAGCCCTTGCCCTTGAATGTTAGTGAATCTGTGCTGCTAGA 232
|||||
51 GlyAsnArgArgArgPheArgValArgGlnProLeuLeuGlnTyrArgTr 67
|||||
233 GGAACCCGTAGCGGTTCCCGTTAGGAGCCCATCTGCGAGTATAGTG 282
|||||
67 pAspIleMetHisArgLeuGluGluProGlnAlaArgMetArgGluGlu 84
|||||
283 GGACATAATGATAGCTTGGAGAGCCAGGCAAGGATGAGAGGAGAGA 332
|||||
84 sMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
|||||
333 ATATGGAAGGATGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGG 382
|||||
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProH1 117
|||||
383 GAAACCCAGTCACTAGTCTCGGGCAGTCAGCACTGATCCCCCTCA 432
|||||
117 sHisAspHisAspGluPheCysLeuMetPro 128
|||||
433 CCATGACCATCAGTATGTTTTCCTTATGCC 466
|||||
seq_name: gb_est1:A1291126

```

seq_documentation_block:
LOCUS A1291126 669 bp mRNA EST 29-JAN-1999
DEFINITION qm15f02.x1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1881915 3', similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74 (HUMAN); mRNA sequence.
ACCESSION A1291126

84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
|||||
333 ATATGGAAGAGTTGGGAGGAGGTGACAGCTGATGGAAGAGCTGAGG 382

101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProH 117
|||||
383 GAAAGACAGTTGAGTCATAGTCTGGGCGCAGTGCACCTGATCCCTCA 432

117 sHisAspHisHisAspGluPheCysLeuMetPro 128
|||||
433 CCATGACCATCAGCATGAGTTGCTTATGCC 466

seq_name: gb_est2:BF237433

seq_documentation_block:
LOCUS BF237433 711 bp mRNA EST 14-NOV-2000
DEFINITION 601842108F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4079856 5',
mrna sequence.
ACCESSION BF237433
VERSION BF237433.1 GI:11151351
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 711)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM938 row: p column: 01
High quality sequence stop: 708.

FEATURES
Location/Qualifiers
1..711
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4079856"
/clone_lib="NIH_MGC_46"
/tissue_type="leiomyosarcoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pOTB7; Site: 1: XhoI; Site: 2:
EcoRI; cDNA made by oligo-dT priming, directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

BASE COUNT 184 a 152 c 218 g 157 t
ORIGIN

alignment_scores:
Quality: 664.00 Length: 128
Ratio: 5.270 Gaps: 0
Percent Similarity: 98.438 Percent Identity: 98.438

alignment_block:
US-09-327-750D-31 x BF237433
Align seg 1/1 to: BF237433 from: 1 to: 711

1 MetGluSerLysGluGluArgAlaLeuAsnLeuIleValGluAsnVa 17

AI291126.1 GI:3933900
EST.
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 669)
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D.; Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbbrp/image/image.html
Insert Length: 749 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 466.
Location/Qualifiers
1..669
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1881915"
/clone_lib="NCI_CGAP_Lu5"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pTT3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
neuroendocrine lung carcinoid, and was then primed with a
Not I - oligo(dT) primer. Double-stranded cDNA was ligated
to Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pTT73 vector. Library is normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 186 a 132 c 185 g 164 t 2 others
ORIGIN

alignment_scores:
Quality: 664.00 Length: 128
Ratio: 5.270 Gaps: 0
Percent Similarity: 98.438 Percent Identity: 98.438

alignment_block:
US-09-327-750D-31 x AI291126
Align seg 1/1 to: AI291126 from: 1 to: 669

1 MetGluSerLysGluGluArgAlaLeuAsnLeuIleValGluAsnVa 17
83 ATGGAGTCCAAAGAGACGACGCGTTAAACAATCTCATCGGGAATGT 132

17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
133 CAACAGGAAATATGATAAAAGATGAAAGGACGAGTTCCTAATAAG 182

34 lylGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
183 GGAGGCCCTTGGCCCTTACCTTTGAATGTAGTGAATACTGTGTCCTAGA 232

51 GlyAsnArgArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
233 GGAACCGTAGCGGTTCGCCCTTAGGAGCCCATCCTGCGAGTAGATG 282

67 pAspIleMetHisArgLeuGluGluProGlnAlaArgMetArgGluGlu 84
283 GGACATAATGCATAGGCTTGGAGAGCCACAGGACGAGGATGAGAGGAGA 332

```

|||||
212 ATGGAGTCCAAAGAGGAACGACGCTTAACAATCTCATCTGGAATAATGT 261
17 lAsnGlnGluAsnAspGluLysAspGluGluGlnValAlaAsnLysG 34
|||||
262 CAACCAAGAAATGATGAAGAAGATGAAGAGGACCAAGTTGCTTAATAAG 311
34 lylGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
312 GGGAGCCCTTGCCCTACCTTTGAATGTTAGTGAATACTGTGTGCCTAGA 361
51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
|||||
362 GGAACCCGTAGCGGTTCCGGTTAGCAGCCCATCTGCAGTATAGTG 411
67 pasPileMetHisArgLeuGluGluProGlnAlaArgMetArgGluGlu 84
|||||
412 GGACATAATGATAGGCTTGAGAGCCACAGGCAAGGATGAGAGAGGAGA 461
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
|||||
462 ATATGGAAGGATTGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGG 511
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 117
|||||
512 GAAAGACAGTTGAGTCATAGTTTCGGGGCAGTCAGCACTGATCCCCCTCA 561
117 shisAspHisAspGluPheCysLeuMetPro 128
|||||
562 CCATGACCATCAGCATGAGTTGCTTATGCCCT 595

```

seq_name: gb_est1:BE266012

```

seq_documentation_block: 610 bp mRNA EST 13-JUL-2000
LOCUS BE266012
DEFINITION 601194731f1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3538735 5',
mRNA sequence.
ACCESSION BE266012
VERSION BE266012.1 GI:9139583
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
1 (bases 1 to 610)
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-remail.nih.gov
Plate: LCM224 row: m column: 08
High quality sequence stop: 607.
FEATURES
source
1..610
Location/Qualifiers
1..610
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3538735"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site: 1: XhoI; Site: 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using 2AP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT
ORIGIN

```

alignment_scores:
Quality: 663.00 Length: 129
Ratio: 5.180 Gaps: 1
Percent Similarity: 99.225 Percent Identity: 98.450
alignment_block:
US-09-327-750D-31 x BE266012
Align seg 1/1 to: BE266012 from: 1 to: 610
1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
|||||
172 ATGGAGTCCAAAGAGGAACGACGCTTAACAATCTCATCTGGAATAATGT 221
17 lAsnGlnGluAsnAspGluLysAspGluGluGlnValAlaAsnLysG 34
|||||
222 CAACCAAGAAATGATGAAGAAGATGAAGAGGACCAAGTTGCTTAATAAG 271
34 lylGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
272 GGGAGCCCTTGCCCTACCTTTGAATGTTAGTGAATACTGTGTGCCTAGA 321
51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
|||||
322 GGAACCCGTAGCGGTTCCGGTTAGCAGCCCATCTGCAGTATAGTG 371
67 pasPileMetHisArgLeuGluGluProGlnAlaArgMetArgGluGlu 84
|||||
372 GGACATAATGATAGGCTTGAGAGCCACAGGCAAGGATGAGAGAGGAGA 421
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuAr 100
|||||
422 ATATGGAAGGATTGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAG 471
100 gLulysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 117
|||||
472 GGAAGACAGTTGAGTCATAGTTTCGGGGCAGTCAGCACTGATCCCCCTC 521
117 shisAspHisAspGluPheCysLeuMetPro 128
|||||
522 ACCATGACCATCAGCATGAGTTTGCCTTATGCCCT 556

```

OM of: US-09-327-750d-31 to: Issued_Patents_NA:* out_format : pfs
 Date: Mar 11, 2002 3:50 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
 -MODEL=framet_p2n.model -DEV=xlh
 -Q/cgn2_1/USPTO_spool/US09327750/runat_11032002_101154_20340/app_query.fasta_1.1472
 -DB=Issued_Patents_NA -OFFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdd -LIST=MIN-0 -DOCLIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USPTO=US09327750 -CGN1_1_99 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-327-750d-31
 Query length: 128
 Database: Issued_Patents_NA*
 Database sequences: 351203
 Database length: 113238999
 Search time (sec): 146.090000

score_list:	sequence	Strd Orig	25score	Escore	Len	Documentation
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-033-18 +		87.00	158.56	0.3508	2004
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-043A-18 +		87.00	158.56	0.3508	2004
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-463-483A-18 +		87.00	158.56	0.3508	2004
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-046A-18 +		87.00	158.56	0.3508	2004
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-470-566B-18 +		87.00	158.56	0.3508	2004
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-469-334-18 +		87.00	158.56	0.3508	2004
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-300-529-18 +		87.00	158.56	0.3508	2004
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-033-35 +		87.00	155.90	0.4929	2576
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-044-35 +		87.00	155.90	0.4929	2576
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-463-483A-35 +		87.00	155.90	0.4929	2576
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-046A-35 +		87.00	155.90	0.4929	2576
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-470-566B-35 +		87.00	155.90	0.4929	2576
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-469-334-35 +		87.00	155.90	0.4929	2576
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-300-529-35 +		87.00	155.90	0.4929	2576
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-033-17 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-033-26 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-044-17 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-044-26 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-463-483A-17 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-463-483A-26 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-046A-17 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-046A-26 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-470-566B-17 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-470-566B-26 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-469-334-17 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-469-334-26 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-300-529-17 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-300-529-26 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-033-49 +		87.00	151.17	0.9042	4031
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-044-49 +		87.00	151.17	0.9042	4031
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-463-483A-49 +		87.00	151.17	0.9042	4031
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-046A-49 +		87.00	151.17	0.9042	4031
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-470-566B-49 +		87.00	151.17	0.9042	4031
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-469-334-49 +		87.00	151.17	0.9042	4031
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-300-529-49 +		87.00	151.17	0.9042	4031
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-147-777-3 +		86.00	129.75	14.12	24979
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-452-872-3 +		86.00	129.75	14.12	24979
	/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-03985-3 +		86.00	129.75	14.12	24979
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-306-691B-23 +		79.00	139.85	3.86	2301
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:PCT-US93-06251-78 +		79.00	139.85	3.86	2301

/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-513-974B-31 - 78.50 146.47 1.65
 /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-776-971-25 - 78.50 146.47 1.65
 /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-513-971B-322 - 78.50 144.55 2.11
 /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-776-971-103 - 78.50 144.55 2.11
 /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-998-416-141 - 78.00 150.69 0.9614

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-033-18

seq documentation block:
 ; Sequence 18, Application US/08471033
 ; Patent No. 5770696
 ; GENERAL INFORMATION:
 ; APPLICANT: Warren, Gregory W
 ; APPLICANT: Kozziel, Michael G
 ; APPLICANT: Mullins, Martha A
 ; APPLICANT: Nye, Gordon J
 ; APPLICANT: Carr, Brian
 ; APPLICANT: Desai, Nalin M
 ; APPLICANT: Kostichka, N. Kristy
 ; APPLICANT: Duck, Nicholas B
 ; APPLICANT: Estruch, Juan J
 ; TITLE OF INVENTION: NO. 5770696el Pesticidal Proteins and Strains
 ; NUMBER OF SEQUENCES: 50
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CIBA-GEIGY Corporation
 ; STREET: 7 Skyline Drive
 ; CITY: Hawthorne
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10532
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30B
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/471,033
 ; FILING DATE:
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/314,594
 ; FILING DATE: 09-SEP-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/218,018
 ; FILING DATE: 23-MAR-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/037,057
 ; FILING DATE: 25-MAR-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Pace, Gary M.
 ; REGISTRATION NUMBER: P-40,403
 ; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 919-541-8582
 ; TELEFAX: 919-541-8689
 ; INFORMATION FOR SEQ ID NO: 18:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2004 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHEICAL: NO
 ; ANTI-SENSE: NO
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: 1..2004
 ; OTHER INFORMATION: /note= "Maize optimized DNA
 ; OTHER INFORMATION: sequence for VIPIA(a) 80 kd protein from AB78"
 ; US-08-471-033-18
 ; alignment_scores:

```

CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: PAGE, Gary M.
REGISTRATION NUMBER: 40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SQLv3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2004 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..2004
OTHER INFORMATION: /note= "Maize optimized DNA
sequence for VIPIA(a) 80 kd protein from AB78"
US-08-471-044-18

alignment_scores:
Quality: 87.00 Length: 126
Ratio: 1.338 Gaps: 6
Percent Similarity: 51.587 Percent Identity: 27.778

alignment_block:
US-09-327-750D-31 x US-08-471-044-18 ..
Align seg 1/1 to: US-08-471-044-18 from: 1 to: 2004

15 GluAsnValIaSnGlnGluAsnAspGluLysAspGluGlnValal 31
||||| : : : ||||| : : : : : : : |||
270 GAACGTGAGCATGGAGAGTGATCCTGAGCCCAACGAGACCTGAGCA 319
: : : : : : : : : : : : : : : : : :
31 aAsnTysGlyGluProLeuAlaLeuProLeuAsnValserGluTyrCysv 48
: : : : : : : : : : : : : : : : : :
320 ACAGCGTGGAGAGCACTCGAGCACCACCTGGAGCTACACCAACAC... 365
: : : : : : : : : : : : : : : : : :
48 alProArgGlyAsnArgArgPheArgValArgGlnProIleLeuGln 64
||||| : : : : : : : ||| ||| : : :
366 .....CGAGGGCCGACCGCTGGAGCCGGCNCCTGGTCCCAAGGGCATCAG 410
: : : : : : : : : : : : : : : : : :
65 TyrArgTrpAsp.....SerHisSerLeuArgAlaValse 68
||| : : :
411 CTTGCGGCTGACGGCTGAACCTACAGCAGCAGACCGTGGCCAGAGT 460
69 ....IleMetHisArgLeuGlyGluPro...GlnAlaArgMetArgGluG 83
: : : ||||| : : : ||| ||| : : : |||
461 GGGCACCACGACCGGCACACCCAGCCAGTTCACACCCGACGCC.... 506
83 luAsnMetGluArgIleGlyGluGluValArgGlnInuMetGluLysLeu 99
||||| : : : : : ||| ||| : : : : :
507 .....CGGCTACCTGAACGCCAACGCTGCGTACCAACAAACGCTGGG 545
100 ArgGluLysGlnLeu.....SerHisSerLeuArgAlaValse 112
||||| : : : ||| ||| : : : |||

```



```

; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 23-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
US-08-471-046A-18

```

```

alignment_scores:
  Quality: 87.00      Length: 126
  Ratio: 1.338      Gaps: 6
  Percent Similarity: 51.587      Percent Identity: 27.778

```

alignment_block:

```

US-09-327-750D-31 x US-08-471-046A-18
Align seg 1/1 to: US-08-471-046A-18 from: 1 to: 2004

15 GluAsnValAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAl 31
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
270 GAACGTGAGCATGCGAGAGGTGATCTCGAGCCCAACGAGACCTGACCA 319
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
31 aAsnLysGluProLeuAlaLeuProLeuAsnValSerGluTyrCysV 48
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
320 ACAGCGTGGAGGACCTCGAGCACCACCACTGGAGCTACACCAACAC... 365
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
48 alProArgLysAsnArgArgPheArgValArgGlnProIleLeuGln 64
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
366 .....CGAGGCGCCACGCTGGAGCGGCGTCCCAAGGCATCATG 410
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
65 TyrArgTrpAsp..... 68
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
411 CTTCGGCGTGGAGTGAACCTACCAGCACGAGCCGTCGGCCAGAGT 460
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
69 ....IleMethisArgLeuGlyGluPro...GlnAlaArgMetArgGlu 83
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
461 GGGGCACCCAGCAGCGCAACACCCAGCAGTTCACACCCGCGCG... 506
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
83 LuAsnMetGluArgIleGlyGluValArgGlnLeuMetGluLeu 99
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
507 .....CGGCTACCTGAACGCAACGTCGCTCAACAACGCTGGG 545
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
100 ArgGluLysGlnLeu.....SerHisLeuArgAlaVal 112

```

```

; .....  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
546 CACCGCGGCATCTACGAGCTGAAGCCACCCACCCAGCTCGTCT... 590
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
112 rThraspProProHisHisAspHisHis 121
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
591 ....GAACAACGACCATCGCCACCAT 614
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-470-566B-18

seq_documentation_block:
; Sequence 18: Application US/08470566B
; Patent No. 5872212
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kosziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5872212el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5872212artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,566B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"

```



```
US-08-470-566B-18
alignment_scores:
  Quality: 87.00      Length: 126
  Ratio: 1.338       Gaps: 6
  Percent Similarity: 51.587      Percent Identity: 27.778

alignment_block:
  US-09-327-750D-31 x US-08-470-566B-18
  Align seg 1/1 to: US-08-470-566B-18 from: 1 to: 2004

15 GluAsnValAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAl 31
270 GAACGTGAGCATGGAGAGGTGATCTCGAGCCCAACGAGCACTGAGCA 319
31 aAsnLysGlyLysGluProLeuAlaLeuProLeuAsnValSerGluTyrCysV 48
320 ACAGCGTGAGAGCCACTCGAGCACTGAGCACTGAGCACTGAGCACTGAGCA 365
48 alProArgGlyAsnArgArgPheArgValArgGlnProLeuGln 64
366 .....CGAGGCGCCAGCGTGGAGCGGCATCGTCCCAAGGCATCAG 410
65 TyrArgTTPAsp..... 68
411 CTTCGGCGTGAGCGTGAACCTACCAGCAGCAGCAGCGTGGCCAGGAGT 460
69 .....lleMethHisArgLeuGlyGluPro...GlnAlaArgMetArgGluG 83
461 GGGGACACGACCGCGGACACCGCCAGTTCACACCGCCAGCGC.... 506
83 luAsnMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeu 99
507 .....CGGTACCTGAAGCCCAACGTCGCTACACCAACGTTGGG 545
100 ArgGluLysGlnLeu.....SerHisSerLeuArgAlaValSe 112
546 CACCGCGCCATCTAGCAGGTGAGCGGACCCACCGACCGTTCGTGCT..... 590
112 rThrAspProHisHisAspHisHis 121
591 ....GAACACACGACACCATCGCCACCAT 614

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:us-08-469-334-18

seq_documentation_block:
; Sequence 18, Application US/08469334
; Patent No. 5990383
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5990383el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
100 ArgGluLysGlnLeu.....SerHisSerLeuArgAlaValise 112
      ::::::::::::::
156 CACCGCGGCCATCTACGACGTGAAGCCACCCACCGCTTCGTGCT..... 590
112 rThrasProHisHisHisHis 121
      ::::::::::::::
591 ....GACACAGCACCATCGCCACCAT 614

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-300-529-18
seq_documentation_block:
; Sequence 18, Application US/09300529
; Patent No. 6066783
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: Genes Encoding Insecticidal Proteins
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6066783artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/300,529
; FILING DATE: TBA
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/469,334
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: S-19506L
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
```

```
FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"
; US-09-300-529-18

alignment_scores:
  Quality: 87.00 Length: 126
  Ratio: 1.338 Gaps: 6
Percent Similarity: 51.587 Percent Identity: 27.778

alignment_block:
US-09-327-750D-31 x US-09-300-529-18

Align seg 1/1 to: US-09-300-529-18 from: 1 to: 2004

15 GluAsnValAsnGlnGluAsnAspGluLysAspGluLysGlnValAl 31
      ::::::::::::::
270 GAACGTGACATGGAGAGGTGATCTTGAGCCCAACGAGAACTGAGCA 319
      ::::::::::::::
31 AsnLysGlyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysV 48
      ::::::::::::::
320 ACAGCGTGAGAGGCACCTCGAGCACCACCACTGGAGCTACACCAAC 365
      ::::::::::::::
48 alProArgGlyAsnArgArgPheArgValArgGlnProLeuLeuGln 64
      ::::::::::::::
366 .....CGAGGGCGCCGCGTGGAGCGCGCATCGTCCCAAGGGCATCAG 410
      ::::::::::::::
65 TyrArgTirPasp..... 68
      ::::::::::::::
411 CTTGCGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 460
      ::::::::::::::
69 .....IleMethHisArgLeuGlyGluPro.....GlnAlaArgMetArgLUC 83
      ::::::::::::::
461 GGGGCACCGACCGCGGCAACACCGAGCGGTTCACACCGCCGCGC.... 506
      ::::::::::::::
83 luAsnMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeu 99
      ::::::::::::::
507 .....CGGCTACCTGACGCGCAACGCTGCGCTACCAACACGTTGG 545
      ::::::::::::::
100 ArgGluLysGlnLeu.....SerHisSerLeuArgAlaValise 112
      ::::::::::::::
546 CACCGCGGCCATCTACGACGTGAAGCCACCCACCGCTTCGTGCT.... 590
      ::::::::::::::
112 rThrasProHisHisHisHisHis 121
      ::::::::::::::
591 ....GACACAGCACCATCGCCACCAT 614

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-033-35
seq_documentation_block:
; Sequence 35, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
```

ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,033
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: P-40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLv3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 2576 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 9...2564
OTHER INFORMATION: /note= "Maize optimized sequence
encoding VIP1A(a) with the Bacillus secretion signal removed
OTHER INFORMATION: contained in pcib5526"
US-08-471-033-35

alignment_scores:
Quality: 87.00 Length: 126
Ratio: 1.338 Gaps: 6
Percent Similarity: 51.587 Percent Identity: 27.778

alignment_block:
US-09-327-750D-31 x US-08-471-033-35 ..
Align seg 1/1 to: US-08-471-033-35 from: 1 to: 2576

15 GluAsnValAsnGlnGluAspGluLysGluGlnValAl 31
833 GAACGTGAGCAGTGGAGGAGTCTCTGAGCCCAACGAGCACTGAGCA 882
31 aAsnLysGlyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysv 48
883 ACAGCTGGAGAGCCACTCGAGCAACCACTGGAGCTACACCAACAC.... 928
48 alProArgGlyAsnArgArgPheArgValArgGlnProIleLeuGln 64
929CGAGGGCCCAAGCGTGGAGCCGCGCATCGGTCCCAAGGCGCATCAG 973
65 TyrArgTrpAsp..... 68
974 CTTCCGGCTGAGGTGAACCTACACAGCAGCAGCAGCGTGGCCAGGAGT 1023
69IleMetHisArgLeuGlyGluPro....GlnAlaArgMetArgGluG 83
1024 GGGGACACGACCGGCAACACACGAGCTTCAACACCGCGCGC.... 1069

83 luAsnMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeu 99
1070CGGCTACCTACGACCGCAACGTCGCTACAAACAGTGGG 1108
100 ArgGluLysGlnLeu.....SerHisSerLeuArgAlaValSe 112
1109 CACGGGGCCATCTACGACGTGAAGCCACACACGAGCTTCGTGCT..... 1153
112 rThrAspProHisHisHisHis 121
1154GAACACGACACCATCGCCACCAT 1177
seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-471-044-35
seq_documentation_block:
Sequence 35, Application US/08471044
Patent No. 5840868
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Koziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,044
FILING DATE: 06-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLv3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 2576 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"

STATE: NY USA
COUNTRY: USA
ZIP: 10532
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,483A

APPLICATION NUMBER: US/08/463,483A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/0037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Sprull, W. Murray
REGISTRATION NUMBER: 37,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
COMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 2576 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
PHETICAL: NO
NAME/KEY: CDS
LOCATION: 9..2564
OTHER INFORMATION:
OTHER INFORMATION: /note= "Maize optimized sequence
encoding VIP1A(a) with the Bacillus
63-483A-35 contained in pCIB5526"

Length:	126
Gaps:	6
Percent Identity:	27.778

ent_block: 327-750D-31 x US-08-463-483A-35

[illegible]

```

1024 GGGCACCAGCACCAGCGCAACACAGCCAGCTTCAACACCGCCAGCG.... 1069
      :::::::::: ::::: ||| :::::||||
83 luAsnMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeu 99
      ::::: ||::: ||| :::::||||:
1070 .....CGGTACTTGAACGCCAACGTGGCTACAAACAACGTGG 1108
      :::::||||:
100 ArgGluLysGlnLeu.....SerHisSerLeuArgAlaValSer 112
      :::::||||:
1109 CACGGGGCGCATCTACGACGTGAAGCCCAACACCGTTCGTGCT..... 1153
      :::::||||:
112 rThrAspProHisHisAspHisHis 121
      :::::|||||
1154 ...GAACAACGACCATTCGCCACCAT 1177
      :::::|||||

```

seq name: /cqn2 6/prodata/2/ina/5B COMB.seq:US-08-471-046A-35

```

seq_documentation_block:
; Sequence 35, Application US/08471046A
; Patent No. 5866326
;
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estuch, Juan J
;
; TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal
; PROTEIN GENES
;
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5866326artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
;
; CURRENT APPLICATION DATA:
; FILING APPLICATION NUMBER: US/08/471,046A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SQLv4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
;
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```



```

APPLICANT: Estruch, Juan J
TITLE OF INVENTION: Genes Encoding Insecticidal Proteins
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSSEE: NO. 606678artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/300,529
FILING DATE: TBA
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/469,334
FILING DATE: 06-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-19506L
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 2576 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 9..2564
OTHER INFORMATION: /note= "Maize optimized sequence encoding VIP1A(a) with the Bacillus secretion signal rema
OTHER INFORMATION: contained in pCIB5526"
OTHER INFORMATION:
US-09-300-529-35

alignment_scores:
Quality: 87.00 Length: 126
Ratio: 1.338 Gaps: 6
Percent Similarity: 51.587 Percent Identity: 27.778

alignment_block:
US-09-327-750D-31 x US-09-300-529-35 ..
Align seg 1/1 to: US-09-300-529-35 from: 1 to: 2576

15 GluAsnValAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAl 31
||||| ::: ||||| ::: ||||| ::: ||||| ::: ||||| :::
833 GACGGTGCATCGAGCAAGGTGATCTCTGAGCCGCCAACGAGAACTGACGA 882

```

```
31 aAsnLysGlyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysv 48
      ::::::::::::::::::::
883 ACAGCGTGAGAGCCACTCGAGCACCACCTACACCAACAC..... 928
      ::::::::::::::::::::
48 alProArgGlyAsnArgArgPheArgValArgGlnProIleLeuGln 64
      ::::::::::::::::::::
929 .....CGAGGGCGCGAGCGTGAGGCGCGCATCGGTCCCAAGGCGCATCAG 973
      ::::::::::::::::::::
65 TyrArgTirPasp..... 68
      ::::::::::::::::::::
974 CTTGCGCGTGAGCGTGAACCTACCAGCAGCAGCAGCGTGCGCCAGGAGT 1023
      ::::::::::::::::::::
69 .....IleMetHisArgLeuGlyGluPro.....GlnAlaArgMetArgGluG 83
      ::::::::::::::::::::
1024 GGGGACACGACCGCCGACACCGCCAGTTCACACCGCCAGCGC..... 1069
      ::::::::::::::::::::
83 luAsnMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeu 99
      ::::::::::::::::::::
1070 .....CGGTACTCGAAGCGCAACGCTGCGCTACAAACACGTTGGG 1108
      ::::::::::::::::::::
100 ArgGluLysGlnLeu.....SerHisSerLeuArgAlaValSe 112
      ::::::::::::::::::::
1109 CACCGCGCCATCTACGAGCTGAAGCCACACCGCCAGCTTCGTGCT..... 1153
      ::::::::::::::::::::
112 rThrAspProProHisHisAspHisHis 121
      ::::::::::::::::::::
1154 .....GAACAACGACACCATCGCCACCAT 1177
      ::::::::::::::::::::
seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-471-033-17

seq_documentation_block:
; Sequence 17, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.308
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,033
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: P-40,403
```

```
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLv3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2655 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEetical: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..2652
OTHER INFORMATION: /note= "Maize optimized DNA
OTHER INFORMATION: sequence for 100 kd VIPIA(a) protein from AB78"
US-08-471-033-17
```

```
alignment_scores:
Quality: 87.00 Length: 126
Ratio: 1.338 Gaps: 6
Percent Similarity: 51.587 Percent Identity: 27.778
alignment_block:
US-09-327-750D-31 x US-08-471-033-17
Align seg 1/1 to: US-08-471-033-17 from: 1 to: 2655
15 GluAsnValAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAl 31
      ::::::::::::::::::::
921 GAACGTGAGCATGGAGAGGTGATCTCTGAGCCCAACGAGAACCTGAGCA 970
      ::::::::::::::::::::
31 aAsnLysGlyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysv 48
      ::::::::::::::::::::
971 ACAGCGTGAGAGCCACTCGAGCACCACCTGGAGCTACACCAACAC..... 1016
      ::::::::::::::::::::
48 alProArgGlyAsnArgArgPheArgValArgGlnProIleLeuGln 64
      ::::::::::::::::::::
1017 .....CGAGGGCGCGAGCGTGAGGCGCGCATCGGTCCCAAGGCGCATCAG 1061
      ::::::::::::::::::::
65 TyrArgTirPasp..... 68
      ::::::::::::::::::::
1062 CTTGCGCGTGAGCGTGAACCTACCAGCAGCAGCAGCGTGCGCCAGGAGT 1111
      ::::::::::::::::::::
69 .....IleMetHisArgLeuGlyGluPro.....GlnAlaArgMetArgGluG 83
      ::::::::::::::::::::
1112 GGGGACACGACCGCCGACACCGCCAGTTCACACCGCCAGCGC..... 1157
      ::::::::::::::::::::
83 luAsnMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeu 99
      ::::::::::::::::::::
1158 .....CGGTACTCGAAGCGCAACGCTGCGCTACACCAACGCTGGG 1196
      ::::::::::::::::::::
100 ArgGluLysGlnLeu.....SerHisSerLeuArgAlaValSe 112
      ::::::::::::::::::::
1197 CACCGCGCCATCTACGAGCTGAAGCCACCGCCAGCTTCGTGCT..... 1241
      ::::::::::::::::::::
112 rThrAspProProHisHisAspHisHis 121
      ::::::::::::::::::::
1242 .....GAACAACGACACCATCGCCACCAT 1265
      ::::::::::::::::::::
```


251 GGAACCGTAGGGGTCGCTTACGAGCCCACTCTGCGATATAGATG 300
 67 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84
 301 GGACATAATGCATAGCTTGGAGAGCCACAGGCAAGCATGAGAGAGAGA 350
 84 sMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 100
 351 ATATGGAAGAGATTGGGAGAGGTGAGACAGCTGATGGAAGAGCTGAGG 400
 101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHI 117
 401 GAAAGACAGCTGAGTCATAGTCTCGGCGAGTCAGCAGCTATCCCTCA 450
 117 sHisAspHisHisAspGluPheCysLeuMetPro 128
 451 CCATGACCATCATGATGAGTCTTTCCTTATGCCCC 484

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAI58581

seq_documentation_block:

ID AAI58581 standard; cDNA; 862 BP.

AC AAI58581;

DF 22-OCT-2001 (first entry)

DE Human polynucleotide SEQ ID NO 784.

KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW myotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

DR WPI; 2001-442253/47.

DR P-PSDB; AAM39425.

PT Novel nucleic acids and polypeptides, useful for treating disorders

PT such as central nervous system injuries -

PS Claim 1; SEQ ID NO 784; 10078pp; English.

CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

SQ Sequence 862 BP; 235 A; 183 C; 238 G; 206 T; 0 other;

alignment_scores

Quality: 679.00 Length: 128

Ratio: 5.305 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-327-750D-31 x AAI58581

Align seg 1/1 to: AAI58581 from: 1 to: 862

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
 226 ATGGAGTCCAAAGAGGAGGAGCGTTAAACAATCTCATCTGGGAAAATGT 275

17 lAsnGlnGluAsnAspGluLysAspGluGluGlnValAlaAsnLysG 34
 276 CAACGAGGAAAATGATGAAAAGATGAAAAGGAGCAAGTTGCTAATAAAG 325

34 lylGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
 326 GGGAGCCCTTGGCCCTACCTTTGAATGTTAGTAATCTGTGTCCTAGA 375

51 GlyAsnArgArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
 376 GGAACCCGTAGGCGGTTCCGGTTAGGAGCCCATCTCGCAGTATAGATG 425

67 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84
 426 GCACATATGTCATAGCTTGGAGAGCCACAGGCAAGGATGAGAGAGAGA 475

84 sMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 100
 476 ATATGGAAGGATTGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGG 525

101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHI 117
 526 GAAAGCAGTTGAGTCATAGTTTCGGGCGAGTCAGCACTGATCCCTCA 575

117 sHisAspHisHisAspGluPheCysLeuMetPro 128

576 CCATGACCATCATGATGAGTCTTTCCTTATGCCCC 609

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAF59611

seq_documentation_block:

ID AAF59611 standard; cDNA; 898 BP.

AC AAF59611;

DF 24-APR-2001 (first entry)

DE Human cell cycle and proliferation protein CCYPR-22 cDNA, SEQ ID NO:76.

KW Cell cycle and proliferation protein; CCYPR; human; agonist;
 KW antagonist; gene therapy; detection; gene therapy;
 KW transgenic animal disease model; immune disorder;
 KW developmental disorder; cell signalling disorder;
 KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;
 KW arteriosclerosis; asthma; allergy; diabetes mellitus;

menstrual cycle disorder; bacterial infection; ss.
Homo sapiens.
WO200107471-A2.
01-FEB-2001.
21-JUL-2000; 2000WO-US19948.
21-JUL-1999; 99US-0145075.
08-SEP-1999; 99US-0153129.
10-NOV-1999; 99US-0164647.
(INCY-) INCYTE GENOMICS INC.
Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman D;
Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;
WPI: 2001-112727/12.
P-PSDB: AAB60474.
Human cell cycle and proliferation proteins and polynucleotides are
used to treat, diagnose and prevent immune, developmental and cell
signaling disorders and cell proliferative disorders including cancer -
Claim 5; Page 181-182; 205pp; English.
Sequences AAF59590-AAF59643 represent cDNAs encoding 54 human
cell cycle and proliferation proteins (CCYPR), AAB60453-AAB60506.
CCYPR and agonists of CCYPR are used to treat diseases or conditions
associated with decreased expression of functional CCYPR, while CCYPR
antagonists are used to treat diseases or conditions associated with
overexpression of functional CCYPR. Monoclonal or polyclonal antibodies
to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or
radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect
compounds e.g., antibodies, oligonucleotides and proteins (receptors)
that specifically bind to CCYPR, and in drug screening methods to
identify compounds that modulate the activity of CCYPR. CCYPR
nucleotides can be used to generate transgenic animal models of human
disease, and can be used in gene therapy in target cells with genetic
abnormalities with respect to the expression of CCYPR for the
treatment or prevention of a disorder associated with CCYPR.
Diseases which can be diagnosed, treated and prevented using CCYPR
proteins, nucleic acids, agonists or antagonists include immune,
developmental and cell signalling disorders, and cell proliferative
disorders including cancer. Specific examples of these disorders
include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
diabetes mellitus, disorders of the menstrual cycle and infections
caused by bacteria.
Sequence 898 BP; 250 A; 186 C; 251 G; 211 T; 0 other;
alignment_scores:
Quality: 679.00 Length: 128
Ratio: 5.305 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-327-750D-31 x AAF59611 ..
Align seg 1/1 to: AAF59611 from: 1 to: 898
1 MetGluSerLySGluGluArgAlaLeuAsnAsnLeuValGluAsnVa 17
|||||
254 ATGGAGTCCAAAGAGAGAGCGTAAACAAATCTCATCGTGGAAATGT 303
17 laaGlnGluAsnAspGluLySGluGluGluValAlaAsnLySg 34
|||||
304 CAACACAGAAATATGTAAGAAAGATCAAGAGAGCAAGTCTAATAAG 353
34 lyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50

|||||
354 GGAGGCCCTGGCCCTACCTTGAATGTTAGTGAATCTGTGTGCTAGTA 403
51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
|||||
404 GGAACCGTAGCGGTTCGGCGTTAGGAGGCCCATCTCTGCAGTAGATAG 453
67 pasPileMetHisArgLeuGluGluProGlnAlaArgMetArgGluGlu 84
|||||
454 GGACATATGTCATAGCTTGGAGACCCACAGCAGGATGAGAGAGGAGA 503
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLySLeuArg 100
|||||
504 ATATGGAAGAGGATGGGGAGGAGGTGAGACAGCTGATGGAAAGCTGAGG 553
101 GluLySGluLeuSerHisSerLeuArgAlaValSerThrAspProProH1 117
|||||
554 GAAACGAGTTCAGTCATAGTCTCGCGGAGTCAGCAGCTGATCCCTCA 603
117 shiAspHisAspGluPheCysLeuMetPro 128
|||||
604 CCATGACCATCAGGATGAGTTTTCCTTATGCC 637
seq name: /SID52/cgdata/geneseq/geneseq/NA2001.DAT:AAI60367
seq_documentation_block:
ID AAI60367 standard; cDNA; 858 BP.
XX
AC AAI60367;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4356.
XX
KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
WPI: 2001-442253/47.
DR P-PSDB; AAM41211.
XX
Novel nucleic acids and polypeptides, useful for treating disorders
such as central nervous system injuries -
XX
Claim 1; SEQ ID NO 4356; 10078pp; English.
XX
The invention relates to human nucleic acids (AAI57798-AAI61369) and

CC the encoded polypeptides (AA38642-AA42213), with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

SQ Sequence 858 BP; 236 A; 175 C; 241 G; 206 T; 0 other;

alignment_scores:
 Quality: 666.00 Length: 129
 Ratio: 5.203 Gaps: 1
 Percent Similarity: 99.225 Percent Identity: 99.225

alignment_block:

US-09-327-750D-31 x AA160367

Align seg 1/1 to: AA160367 from: 1 to: 858

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
 |||||
 215 ATGGAGTCCAAAGAGGACGAGGCTTAACAATCTCATCTGCGAAATGT 264
 17 IasnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
 |||||
 265 CAACACGAGAAATGATCAAAAGATGAAAGGAGCAAGTTGCTATAAAG 314
 34 IyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
 |||||
 315 GGGAGCCCTTGGCCCTACCTTTGAATGTTAGTAATCTGTGTGCTAGA 364
 51 GlyAsnArgArgPheArgValArgGlnProIleLeuGluTyrArgTr 67
 |||||
 365 GGAACCCGTAGCGGTTCCGCTTAGCGAGCCCATCTGCAGTATAGTG 414
 67 PaspIleMethHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84
 |||||
 415 GGACATATGATAGCTAGGCTTGAGAGCCACAGGACAGGATGAGAGAGGA 464
 84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
 |||||
 465 ATATGGAAGGATTGGGAGGAGGTGAGACAGCTGATGGAAGCTGAGG 514
 101 GlnLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProH 117
 |||||
 515 GAAAGAGCTTGATGATAGTCTGCGGGCAGTCAGCAGCTGATCCCTCA 564
 117 shiAspHisAspGluPheCys LeuMetPro 128
 |||||
 565 CCATGACCATCAGCATGAGTTTGGCCCTTATGCC 599

seq_name: /SDS2/gcgdata/geneseq/geneseq/NA2000.DAT: AAC03880

seq_documentation_block:

ID AAC03880 standard; cDNA; 662 BP.

XX AC AAC03880;

XX AC AAC03880;

DT 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 3878.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

gene therapy; chromosome mapping; ss.
 XX Homo sapiens.
 XX EP1033401-A2.
 XX 06-SEP-2000.
 PF 21-FEB-2000; 2000EP-0200610.
 PR 26-FEB-1999; 99US-0122487.
 PA (GEST) GENSET.
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 XX WPI; 2000-500381/45.
 DR P-PSDB; AAG03874.
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX Claim 1; SEQ ID 3878; 71pp + CD-ROM; English.
 XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.
 SQ Sequence 662 BP; 177 A; 144 C; 192 G; 137 T; 12 other;

alignment_scores:

Quality: 649.00 Length: 128
 Ratio: 5.234 Gaps: 0
 Percent Similarity: 96.875 Percent Identity: 95.312

alignment_block:

US-09-327-750D-31 x AAC03880

Align seg 1/1 to: AAC03880 from: 1 to: 662

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
 |||||
 207 ATGGAGTCCAAAGAGGACGAGGCTTAACAATCTCATCTGCGAAATGT 256
 17 IasnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
 |||||
 257 CAACACGAGAAATGATGAAAGATGAAAGGAGCAAGTTGCTATAAAG 306
 34 IyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
 |||||
 307 GGGAGCCCTTGGCCCTACCTTTGATGTTGTGAATACTGTGTGCTAGA 356
 51 GlyAsnArgArgPheArgValArgGlnProIleLeuGluTyrArgTr 67
 |||||
 357 GGAATCGTAGCGGTTCCGCTTAGCAGCCCATCTGCTGAGTATAGTG 406
 67 PaspIleMethHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84
 |||||
 407 GGATATGATGATAGCTAGGCTTGAGAGAACACAGGACAGGATGAGAGAGA 456

84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
 457 ATATGGAAGGATTGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGG 506
 101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH1 117
 507 GAAAGCAGTTGAGTCATAGTCTGGGGCAGCTCAGCAGTACCCCTCA 556
 117 sHisAspHisAspGluPheCysLeuMetPro 128
 557 CCATGACCATCATGATGAGTTTGCNNWATGCC 590

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.AAC03879

seq_documentation_block:

ID AAC03879 standard; cDNA: 698 BP.

AC AAC03879;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 3877.

DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX P-PSDB; AAG03873.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 3877; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.

XX Sequence 698 BP; 187 A; 151 C; 210 G; 139 T; 11 other;

alignment_scores:

Quality:	649.00	Length:	128
Ratio:	5.234	Gaps:	0
Percent Similarity:	96.875	Percent Identity:	95.312

alignment_block:

US-09-327-750D-31 x AAC03879
 Align seg 1/1 to: AAC03879 from: 1 to: 698
 1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
 243 ATGAGTCCAAAGAGGAGGAGCGTTAAACAATCTCATCGTGGAATGT 292
 17 lasGlnGluAsnAspGluLysAspGluGlnValAlaAsnLysG 34
 293 CAACACGAGAAATGATGAAAAAGATGAAAGGAGCAGTGTGCTAATAAG 342
 34 lyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
 343 GGGAGCCCTGGCCCTACCTTTGRATGVTGTAATACTGTGTGCTAGA 392
 51 GlyAsnArgArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
 393 GGAATCGTAGCGGTTCCCGCTTAGGCAGCCCATCTCCGACGTATAGATG 442
 67 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84
 443 GGATATGATCCATAGCTTGGAGAACACAGCAGGATGAGAGAGAGA 492
 84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
 493 ATATGGAAGGATTGGGAGGAGGTGAGACAGCTGATGGAAGCTGAGG 542
 101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH1 117
 543 GAAAGCAGTTGATCATAGTCTGGGCGGAGTGCAGCACTGACCCCTCA 592
 117 sHisAspHisAspGluPheCysLeuMetPro 128
 593 CCATGACCATCATGATGAGTTTGCNNWATGCC 626

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.AAC10889

seq_documentation_block:

ID AAC10889 standard; cDNA: 692 BP.

XX AAC10889;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 14964.

DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 14964; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from

117 SHIsAspHisHisAspGluPheCysLeuMetPro 128
|||||
514 CCATGACCATCATGATGAGTTTGGCTTATGCC 547
seq_name: /SID52/gcgdata/geneseq/NA2000.DAT:AAC06183
seq_documentation_block:
ID AAC06183 standard; cDNA: 421 BP.
AC AAC06183;
XX
XX 06-OCT-2000 (first entry)
XX Human secreted protein 5' EST, SEQ ID NO: 10258.
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
OS Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1: SEQ ID 10258; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX Sequence 421 BP; 110 A; 103 C; 124 G; 82 T; 2 other;

alignment_scores:
Quality: 284.00 Length: 72
Ratio: 4.303 Gaps: 1
Percent Similarity: 91.667 Percent Identity: 84.722
alignment_block:
US-09-327-750D-31 x AAC06183
Align seg 1/1 to: AAC06183 from: 1 to: 421
1 MetGluSerLysGluGluAlaLeuAsnValLeuValGluAsnVa 17
|||||
207 ATGGAGTCCAAAGAGGACGAGGCTTAACAATCTCATCGTGGAAATGT 256

17 lAsnGlnGluAsnAspGluLysAspGluGlnValAlaAsnLysG 34
|||||
257 CAACCCAGGAAATGATGAAGAAGATGAAGAAGCAAGTTCCTAATAAG 306
34 lylGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
307 GGGAGGCCCTTGGCCCTACCTTTGAATGTTAGTGAATACTGTGTGCCTAGA 356
51 GlyAsnArgArgPheArgValArgGlnProIleLeuGln.TyrArgT 67
|||||
357 GGAACCCCTAG.CGGTCCGGCTTAGGCACACAGTATTAGAGGCCCGCT 405
67 rpAspIleMethHis 71
::: |||
406 GCCCAGTGACAT 419
seq_name: /SID52/gcgdata/geneseq/NA2000.DAT:AAC06180
seq_documentation_block:
ID AAC06180 standard; cDNA: 457 BP.
XX
XX AAC06180;
XX
XX 06-OCT-2000 (first entry)
XX Human secreted protein 5' EST, SEQ ID NO: 10255.
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1: SEQ ID 10255; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX Sequence 457 BP; 120 A; 110 C; 142 G; 84 T; 1 other;

alignment_scores:
Quality: 284.00 Length: 72
Ratio: 4.303 Gaps: 1
Percent Similarity: 91.667 Percent Identity: 84.722
alignment_block:
US-09-327-750D-31 x AAC06183
Align seg 1/1 to: AAC06183 from: 1 to: 421
1 MetGluSerLysGluGluAlaLeuAsnValLeuValGluAsnVa 17
|||||
207 ATGGAGTCCAAAGAGGACGAGGCTTAACAATCTCATCGTGGAAATGT 256

117 SHIsAspHisHisAspGluPheCysLeuMetPro 128
|||||
514 CCATGACCATCATGATGAGTTTGGCTTATGCC 547
seq_name: /SID52/gcgdata/geneseq/NA2000.DAT:AAC06183
seq_documentation_block:
ID AAC06183 standard; cDNA: 421 BP.
AC AAC06183;
XX
XX 06-OCT-2000 (first entry)
XX Human secreted protein 5' EST, SEQ ID NO: 10258.
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
OS Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1: SEQ ID 10258; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX Sequence 421 BP; 110 A; 103 C; 124 G; 82 T; 2 other;

alignment_scores:
Quality: 284.00 Length: 72
Ratio: 4.303 Gaps: 1
Percent Similarity: 91.667 Percent Identity: 84.722
alignment_block:
US-09-327-750D-31 x AAC06183
Align seg 1/1 to: AAC06183 from: 1 to: 421
1 MetGluSerLysGluGluAlaLeuAsnValLeuValGluAsnVa 17
|||||
207 ATGGAGTCCAAAGAGGACGAGGCTTAACAATCTCATCGTGGAAATGT 256

117 SHIsAspHisHisAspGluPheCysLeuMetPro 128
|||||
514 CCATGACCATCATGATGAGTTTGGCTTATGCC 547
seq_name: /SID52/gcgdata/geneseq/NA2000.DAT:AAC06183
seq_documentation_block:
ID AAC06183 standard; cDNA: 421 BP.
AC AAC06183;
XX
XX 06-OCT-2000 (first entry)
XX Human secreted protein 5' EST, SEQ ID NO: 10258.
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
OS Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1: SEQ ID 10258; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX Sequence 421 BP; 110 A; 103 C; 124 G; 82 T; 2 other;

alignment_scores:
Quality: 284.00 Length: 72
Ratio: 4.303 Gaps: 1
Percent Similarity: 91.667 Percent Identity: 84.722
alignment_block:
US-09-327-750D-31 x AAC06183
Align seg 1/1 to: AAC06183 from: 1 to: 421
1 MetGluSerLysGluGluAlaLeuAsnValLeuValGluAsnVa 17
|||||
207 ATGGAGTCCAAAGAGGACGAGGCTTAACAATCTCATCGTGGAAATGT 256

117 SHIsAspHisHisAspGluPheCysLeuMetPro 128
|||||
514 CCATGACCATCATGATGAGTTTGGCTTATGCC 547
seq_name: /SID52/gcgdata/geneseq/NA2000.DAT:AAC06183
seq_documentation_block:
ID AAC06183 standard; cDNA: 421 BP.
AC AAC06183;
XX
XX 06-OCT-2000 (first entry)
XX Human secreted protein 5' EST, SEQ ID NO: 10258.
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
OS Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1: SEQ ID 10258; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX Sequence 421 BP; 110 A; 103 C; 124 G; 82 T; 2 other;

alignment_scores:
Quality: 284.00 Length: 72
Ratio: 4.303 Gaps: 1
Percent Similarity: 91.667 Percent Identity: 84.722
alignment_block:
US-09-327-750D-31 x AAC06183
Align seg 1/1 to: AAC06183 from: 1 to: 421
1 MetGluSerLysGluGluAlaLeuAsnValLeuValGluAsnVa 17
|||||
207 ATGGAGTCCAAAGAGGACGAGGCTTAACAATCTCATCGTGGAAATGT 256

Percent Similarity: 91.667 Percent Identity: 84.722

alignment_block:

US-09-327-750D-31 x AAC06180

Align seg 1/1 to: AAC06180 from: 1 to: 457

1 MetGluSerLysGluAlaLeuAsnLeuValGluAsnVa 17
|||||
243 ATGGAGTCCAAAGAGGACGAGCTTAAACAATCTCATCGTGAATAATGT 292
17 lasnGlnGluAsnAspGluLysAspGluGlnValAlaAsnLysG 34
|||||
293 CACACAGGAATGATGAAAGATGAAAGGAGCAAGTTGCTAATAAAG 342
34 lyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
343 GGGAGCCCTTGGCCCTACCTTTGAATGTTAGTGAATACTGTGTGCTAGA 392
51 GlyAsnArgArgGluPheArgValArgGlnProIleLeuGln.TyrArgT 67
|||||
393 GGAACCGTAG.CGTTCCGCTTAGGCACACAGTATTAGAGCCCGCT 441
67 rpAspIleMethis 71
::: |||
442 GCCCAGTGACACAT 455

seq_name: /SID2/gcgdata/geneseq/geneseq/NA2000.DAT:AAC06184

seq_documentation_block:

ID AAC06184 standard; cDNA; 451 BP.

AC AAC06184;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 10259.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GEST.) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI; 2000-500381/45.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

PS Claim 1; SEQ ID 10259; 71pp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.

XX Sequence 451 BP; 117 A; 110 C; 131 G; 91 T; 2 other;

alignment_scores:

Quality: 261.00 Length: 69
Ratio: 4.143 Gaps: 1
Percent Similarity: 91.304 Percent Identity: 81.159

alignment_block:

US-09-327-750D-31 x AAC06184

Align seg 1/1 to: AAC06184 from: 1 to: 451

4 LysGluGluArgAlaLeuAsnLeuValGluAsnValAsnGlnG1 20
|||||
246 CAAAGAGAACGAGCGTTAAACAATCTCATCGTGAATAATGTCAACACGGA 295
20 uAsnAspGluLysAspGluLysGluGlnValAlaAsnLysGlyGluProL 37
|||||
296 AAATGATGAAAGATGAAAGGAGCAAGTTGCTAATAAAGGGGAGCCCT 345
37 euAlaLeuProLeuAsnValSerGluTyrCysValProArgGlyAsnArg 53
|||||
346 TGGCCCTACCTTTGAATGTTAGTGAATACTGTGTGCTAGAGGAACCGT 395
54 ArgArgPheArgValArgGlnProIleLeuGln.TyrArgTtpAspIleM 70
|||||
396 AG.CGTTCCGCTTAGGCACACAGTATTAGAGGACCGCTGCCCACTGA 444
70 etHis 71
|||
445 CACAT 449

seq_name: /SID2/gcgdata/geneseq/geneseq/NA2001.DAT:AAF23529

seq_documentation_block:

ID AAF23529 standard; DNA; 891 BP.

AC AAF23529;

DT 22-MAR-2001 (first entry)

DE Human NADE DNA.

KW Neurotrophin receptor; p75-NTR; NGF-induced apoptosis;
KW neurogenetic disease; NF-kappaB; ds.

OS Homo sapiens.

PN WO200075278-A2.

PD 14-DEC-2000.

PF 07-JUN-2000; 2000WO-US15621.

PR 07-JUN-1999; 99US-0327750.

PA (UYCO) UNIV COLUMBIA NEW YORK.

PI Sato T;

DR WPI; 2001-061707/07.

PT New p75-neurotrophin receptor-associated cell death executor (NADE) and
PT the gene encoding NADE, useful for modulating the activity of p75NTR
PT and for detecting neurodegenerative diseases.

PS Disclosure; Fig 1; 134pp; English.

332 GAAATATCAGCGGGGCGAGTTAGCGACTTGTCCCTAATTTTCGATG 381
 67 pAspIleMetHisArg...LeuGlyGluProGlnAlaArgMetArgGluG 83
 382 GGCCATACCTATAGGCATATGAGCACAATGAAGCGAGA.....G 422
 83 luAsnMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeu 99
 423 ATGATGTAGAAAGGTTTGTAGGCAGATGATGGAATCAAGAGAAAGACT 472
 100 ArgGluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
 473 AGGGAACAGCAGATGAGCATATATGCGCTTCCAAACTCCTGAACCT.. 520
 116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128
 521GACAACCATTTATGACTTTTGCCTCATACCT 550

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH13750

seq_documentation_block:

ID AAH13750 standard; cDNA; 1229 BP.

AAH13750;

26-JUN-2001 (first entry)

Human cDNA sequence SEQ ID NO:10656.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

Human sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0300253.

11-JAN-2000; 2000JP-0118776.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID 10656; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

XX
 SQ Sequence 1229 BP; 351 A; 243 C; 296 G; 339 T; 0 other;

alignment_scores:

Quality: 187.00 Length: 129

Ratio: 2.226 Gaps: 4

Percent Similarity: 65.116 Percent Identity: 37.984

alignment_block:

US-09-327-750D_r31 x AAH13750 ..

Align seg 1/1 to: AAH13750 from: 1 to: 1229

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17

191 ATGGAGTCCAAAGAGGAAGTACGCGCAACAATCTCAACGGGGAATGC 240

17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34

241 CCAACAAGAAAC.....GAAGGAGGGGAGCAGGCCGCCACGAGA 281

34 lYcGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50

282 ATGAGAGAAGATCCCGCCATTTGGAGGGGGTGAAGGCCAGAACCTTGA 331

51 GlyAsnArgArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67

332 GGAAATATCAGCGGGGCGAGTTAGCGACTTGTCCCTAATTTTCGATG 381

67 pAspIleMetHisArg...LeuGlyGluProGlnAlaArgMetArgGluG 83

382 GGCCATACCTATAGGCATATGAGCACAATGAAGCGAGA.....G 422

83 luAsnMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeu 99

423 ATGATGTAGAAAGGTTTGTAGGCAGATGATGGAATCAAGAGAAAGACT 472

100 ArgGluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116

473 AGGGAACAGCAGATGAGCATATATGCGCTTCCAAACTCCTGAACCT.. 520

116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128

521GACAACCATTTATGACTTTTGCCTCATACCT 550

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH13750

seq_documentation_block:

ID AAC85548 standard; cDNA; 1364 BP.

AAAC85548;

04-JUN-2001 (first entry)

cDNA encoding CDIFF-4, Incyte ID-No. 1990956CB1.

Human; cell differentiation; CDIFF; agonist; antagonist; epilepsy; cell proliferation; Alzheimer's disease; schizophrenic disorder; arteriosclerosis; cancer; atherosclerosis; diabetes mellitus; ss.

Homo sapiens.

Key Location/Qualifiers

FT CDS 230..612

441 GGCCATACCTAATAGGCATATTAGCACCAATGAAGCGAGA.....G 481
83 luAsnMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeu 99
482 ATGATGTAGAAAGCTTTTGTAGGCAGATGATGTAATCAAGAGAAGACT 531
100 ArgGluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
532 AGGACAGCAGATGAGGAGCTATATGCGCTTCCAACTCTGAACCT.. 579
116 OHHisHisAspHisHisAspGluPheCysLeuMetPro 128
580GACACCATTATGACTTTTGCTCATACCT 609

FT /*tag= a
FT /product= "CDIFF-4"
XX
XX WO200119860-A2.
XX
XX 22-MAR-2001.
XX
XX 14-SEP-2000; 2000WO-US25435.
XX
XX 15-SEP-1999; 99US-0154140.
XX 06-DEC-1999; 99US-0169155.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Tang Yt, Hillman JL, Yue H, Reddy R, Lal P, Shah P, Azimzai Y;
XX Baughn MR, Lu DAM, Bandman O, Shih LL, Patterson C;
XX
XX WPI: 2001-211447/21.
XX P-P5DB: AAB47126.

Isolated polypeptides and polynucleotides involved in cell differentiation are used for treatment, prevention and diagnosis of cell proliferative, developmental and neurological disorders e.g. cancer and Alzheimer's disease.
Claim 5; Page 121; 137pp; English.

The sequences given in AAC85545-72 encode human polypeptides involved in cell differentiation (CDIFF). CDIFF polypeptides and agonists of these are used to treat a disease or condition associated with decreased expression of functional CDIFF. An antagonist of CDIFF is used to treat a disease or condition associated with over expression of functional CDIFF. CDIFF polypeptides may be used for the treatment, prevention and diagnosis of cell proliferative, developmental and neurological disorders, such as Alzheimer's disease, schizophrenia and epilepsy. The CDIFF-4 sequence is homologous to Mus musculus REX-3. This sequence maps to chromosome 1 within the interval from 152.2 to 157.4 centimorgans, to chromosome 3 within the interval from 157.4 to 158.0 centimorgans, and to the X chromosome within the interval from 104.9 to 150.3 centimorgans.

Sequence 1364 BP; 411 A; 269 C; 322 G; 362 T; 0 other;

alignment_scores:
Quality: 187.00 Length: 129
Ratio: 2.226 Gaps: 4
Percent Similarity: 65.116 Percent Identity: 37.984
alignment_block:
US-09-327-750D-31 x AAC85548 ..
Align seg 1/1 to: AAC85548 from: 1 to: 1364
1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
250 ATGGAGTCCAAAGAGAACTAGCGCAACAATCTCAACGGGGGAAATGC 299
17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
300 CCAACAAGAAAC.....GAAGAGGGGAGCAGGCCGCCAGCAGA 340
34 lYGLuProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
341 ATGAAGAAGAAATCCCGCATTTGGAGGGGTGAAGGCCAGACCTGGA 390
51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTyr 67
391 GGAATATCAGCGGGGGGAGTGTAGCGACTTGTCCCTAATTTTCGATG 440
67 pAspIleMetHisArg...LeuGlyGluProGlnAlaArgMetArgGluG 83

351 ATATGGAAGGATTGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGG 400
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProH1 117
|||||
401 GAAAGCAGATTGAGTTCATAGTCTCGGGCAGTCTGAGCAGTATCCCCCTCA 450
117 shIsAspHisLeuSerHisSerLeuMetPro 128
|||||
451 CCATGACCATCAGATGAGTTTGCCTTATGCC 484

seq_name: gb_pat:AX078272

seq_documentation_block: 898 bp DNA PAT 22-FEB-2001
LOCUS AX078272
DEFINITION Sequence 76 from Patent WO0107471.
ACCESSION AX078272
VERSION AX078272.1 GI:13157963
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Hillman,J.L., Lal,P., Tang,Y.T., Yue,H., Au-Young,J., Bandman,O.,
Arimai,Y., Yang,J., Lu,D.A., Baughn,M.R., Patterson,C. and Shah,P.
TITLE Cell cycle and proliferation proteins
JOURNAL Patent: WO 0107471-A 76 01-FEB-2001;
Incyte Genomics, Inc. (US)

FEATURES
source
1..898
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 5664154CB1"

BASE COUNT 250 a 186 c 251 g 211 t
ORIGIN
|||||

alignment_scores:
Quality: 679.00 Length: 128
Ratio: 5.305 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-327-750D-31 x AX078272

Align seg 1/1 to: AX078272 from: 1 to: 898

1 MetGluSerLysGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
|||||
254 ATGGAGTCCAAAGAGGAGGAGGCTTAACAATCTCATGCTGGAATAATGT 303
17 LasnGlnGluAsnAspGluLysAspGluGlnValAlaAsnLysG 34
|||||
304 CAACCCGGAATAATGATGAAGAAGATGAAGAAGGAGCAAGTGTCTAATAAG 353
34 lylGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
354 GGGAGCCCTTGGCCCTACCTTGAATGTAGTGAATACCTGTGCTGCTAGA 403
51 GlyAsnArgArgGluArgGluValArgGlnProIleLeuGlnTyrArgTr 67
|||||
404 GGAACCCGTAGCGGTTCCCGGTAGGAGCCCATCTCTGCATATAGATG 453
67 pAspIleMetHisArgLeuGluProGlnAlaArgMetArgGluGluA 84
|||||
454 GGACATAATGATAGGCTTGAGAGGCCACAGGCAAGGATGAGAGGAGA 503
84 snMetGluArgGileGluGluValArgGlnLeuMetGluLysLeuArg 100
|||||
504 ATATGGAAGGATTGGGAGGAGGTGAGACAGCTGTATGGAAGAGCTGAGG 553
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProH1 117
|||||

554 GAAAGCAGATTGAGTCTGCGGCGCAGCTCAGACACTGATCCCCCTCA 603

117 shIsAspHisLeuSerHisSerLeuMetPro 128
|||||
604 CCATGACCATCAGATGAGTTTGCCTTATGCC 637

seq_name: gb_pr:HSV870H8

seq_documentation_block:
LOCUS HSV870H8 31321 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from cosmid:V870H8, between markers DXS366 and
DXS87 on chromosome X contains ESTs.

ACCESSION 270233
VERSION 270233.1 GI:1235542
KEYWORDS X.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Whiteley,M.
TITLE Direct Submission
JOURNAL Submitted (19-MAR-1995) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, UK. E-mail enquiries: humquerry@sanger.ac.uk

COMMENT
IMPORTANT: This sequence is not the entire insert of clone V870H8.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone V870H8 is at 1 in this sequence. The
true left end of clone V693A8 is at 31221.
V870H8 is from the human chromosome X-specific cosmid library.

FEATURES
source
1..31321
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone="CHC-870H8"
/clone_lib="SCCV"
475..701
/note="L1 element fragment"
802..988
/note="L1 element fragment"
2533..4138
/note="L1 element fragment"
/note="match: multiple ESTs"
4189..4246
/note="29 copies of 2 mer 91 % conserved"
7496..7874
/note="L1 element fragment"
8022..8108
/note="L1 element fragment"
8302..8493
/note="L1 element fragment"
8385..8671
/note="L1 element fragment"
8769..9010
/note="L1 element fragment"
9259..9384
/note="L1 element fragment"
9624..9731
/note="MLTIC element fragment"
9781..9948
/note="MLTIC element fragment"
9890..9948
/note="MLTIB element fragment"
10418..10592
/partial
/note="Alu repeat: matches 298..108 of consensus"
10593..10688
/partial
/note="Alu repeat: matches 96..1 of consensus"
10698..10781
/note="L1 element fragment"

```
repeat_region 11112..11403
/note="Alu repeat: matches 1..308 of consensus"
repeat_region 11922..12004
/note="L1 element fragment"
repeat_region 12062..12308
/note="L1 element fragment"
repeat_region 12423..12707
/partial
/note="Alu repeat: matches 302..1 of consensus"
repeat_region 12710..12916
/note="L1 element fragment"
repeat_region 12979..13081
/note="L1 element fragment"
repeat_region 13473..13532
/note="MLT2A1 element fragment"
repeat_region 15871..15928
/note="L1 element fragment"
repeat_region 18180..18417
/note="L1 element fragment"
repeat_region 18778..19056
/note="Alu repeat: matches 1..308 of consensus"
repeat_region 22279..22350
/note="L1 element fragment"
repeat_region 22449..22995
/note="L1 element fragment"
repeat_region 23239..23374
/note="L1 element fragment"
repeat_region 23371..23606
/note="MLT1B element fragment"
repeat_region 23395..23541
/note="MLT1A element fragment"
repeat_region 23705..23778
/note="MLT1A element fragment"
repeat_region 23805..23960
/note="L1 element fragment"
repeat_region 24011..24541
/note="L1 element fragment"
repeat_region 24638..24770
/partial
/note="Alu repeat: matches 150..1 of consensus"
repeat_region 24829..30975
/note="L1 element fragment"
BASE COUNT 9462 a 6147 c 6129 g 9583 t
ORIGIN

alignment_scores:
  Quality: 679.00      Length: 128
  Ratio: 5.305        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-327-750D-31 x HSV870H8

Align seg 1/1 to: HSV870H8 from: 1 to: 31321

1 MetGluserylGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
|||||
3508 ATGGAGTCCAAAGAGGAGGAGCGTTAAACAATCTCATGCTGGGAAATGT 3557
|||||

17 lasnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
|||||
3558 CAACCCAGAAATGATGAAAAAGATCAAAAGGACGAGTTGCTATAAAG 3607
|||||

34 lyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
3608 GGGAGCCCTTGGCCCTACCTTGAATGTTAGTCAATCTGCTGCTCTAGA 3657
|||||

51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
|||||
3658 GGAACCCGTAGGCGGTCCCGTTAGGACGCCCATCTGTCAGTATAGATG 3707
|||||

67 pAspIleMethHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84
```

```
|||||
3708 GGACATATGCTAGGCTTGGAGAGCCAGGCAAGGATGAGAGAGAGA 3757
|||||
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
|||||
3758 ATATGGAAGGATTGGGAGGAGGTGAGACAGCTGATGGAAAGCTGAGG 3807
|||||
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
|||||
3808 GAAAACAGCTGAGTCATAGTCTCGGGCAGCTCAGCACTGATCCCCCTCA 3857
|||||
117 shisAspHisAspGluPheCysLeuMetPro 128
|||||
3858 CCATGACCATCAGCATGAGTTTTCCTTATGCTCC 3891
seq_name: gb_pr.AL133348

seq_documentation_block:
LOCUS AL133348 40584 bp DNA PRI 06-MAR-2000
DEFINITION Human DNA sequence from clone RPL-79P11 on chromosome Xq21.32-22.1.
Contains the gene for a novel protein similar to mouse Bex2
(brain-expressed X-linked protein 2), ESTs, STSS, GSSs and a
putative CpG island, complete sequence.
ACCESSION AL133348
VERSION AL133348.8 GI:7076386
KEYWORDS HTG; Bex2; CpG island.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 40584)
AUTHORS Wilton,S.
TITLE Direct Submission
JOURNAL Submitted (02-MAR-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Feb 24, 2000 this sequence version replaced gi:6997869.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
RPL-79P11 is from the library RPCR-1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VICTOR: pcypac2
IMPORTANT: This sequence is not the entire insert of clone
RPL-79P11 it may be shorter because we only sequence overlapping
sections once, or longer because we arrange for a small overlap
between neighbouring submissions.
The true left end of clone LLOXNC01-105G4 is at 40485 in this
sequence. The true right end of clone LLOXNC01-177E8 is at 100 in
this sequence.
FEATURES
Location/Qualifiers
1..40584
/organism="Homo sapiens"
/db_xref="taxon:9606"
```



```
alignment_scores:
  Quality: 679.00      Length: 128
  Ratio: 5.305         Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-327-750D-31 x AL133348/rev ..

Align seg 1/1 to reverse of: AL133348 from: 1 to: 40584

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuValGluAsnVa 17
31132 ATGGAGTCCAAAGAGGAGGCGTTAAACAATCTCATCGTGGAAATGT 31083

17 lAsnGlnGluAsnAspGluLysAspGluGlnValAlaAsnLysG 34
31082 CAACCCAGGAAATGATGAAAGAGATGAAAGAGCAAGTCTGCTAATAAG 31033

34 lyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
31032 GGGAGCCCTTGGCCCTACCTTGAATGTTAGTGAATCTGTGCTCTAGA 30983

51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
30982 GGAACCGTAGCGGTTCCGGTTAGGAGGCCATCTGCGAGTATAGATG 30933

67 pAspIleMetHisArgLeuGluGluProGlnAlaArgMetArgGluGluA 84
30932 GGACATAATGATGCTTGGAGAGCCACAGGCAAGGATGAGAGAGGAGA 30883

84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
30882 ATATGGAAGGATTTGGGAGGAGTGAGACAGCTGATGGAAGAGCTGAGG 30833

101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 117
30832 GAAAAGCATGACATGATGAGTTTGGCGGAGTGCAGCTGATCCCTCTCA 30783

117 sHisAspHisAspGluPheCysLeuMetPro 128
30782 CCATGACCATCAGATGAGTTTGGCGGAGTGCAGCTGATCCCTCTCA 30749

seq_name: gb_htg:HSU80B1

seq_documentation_block:
LOCUS      HSU80B1 41029 bp  DNA  HTG  10-JUL-2001
DEFINITION Homo sapiens chromosome X clone LLOXNC01-80B1 map q22.1-22.3, ***
ACCESSION AL022169
VERSION    AL022169.3 GI:13276704
KEYWORDS   HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 41029)
AUTHORS   Bird,C.
JOURNAL   Direct Submission
          Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
          CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
          requests: clonerequest@sanger.ac.uk
          On Mar 12, 2001 this sequence version replaced gi:12750927.
COMMENT   ----- Genome Center
          Center: Sanger Centre
          Center code: SC
          Web site: http://www.sanger.ac.uk
          Contact: humquery@sanger.ac.uk
          ----- Project Information
          Center project name: cu80B1
          ----- Summary Statistics
          Assembly program: XGAP4; version 4.5
          Sequencing vector: plasmid; 108752; 100% of reads
          Chemistry: Dye-terminator ABI; 19% of reads

Chemistry: Dye-terminator Big Dye; 80% of reads
Consensus quality: 41009 bases at least Q40
Consensus quality: 41023 bases at least Q30
Consensus quality: 41027 bases at least Q20
Insert size: 41029; sum-of-ctigs
Insert size: 51751; 0.6% error; agarose-fp
Quality coverage: 9.67x in Q20 bases; sum-of-ctigs Quality
coverage: 7.67x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES             Location/Qualifiers
     source            1..41029
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q22.1-22.3"
                     /clone="LLOXNC01-80B1"
                     /clone_lib="LLOXNC01"
     misc_feature      1..41029
                     /note="assembly_fragment:00890"
BASE COUNT  13249 a 7796 c 7781 g 12203 t
ORIGIN

alignment_scores:
  Quality: 679.00      Length: 128
  Ratio: 5.305         Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-327-750D-31 x HSU80B1 ..

Align seg 1/1 to: HSU80B1 from: 1 to: 41029

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuValGluAsnVa 17
18412 ATGGAGTCCAAAGAGGAGGCGTTAAACAATCTCATCGTGGAAATGT 18461

17 lAsnGlnGluAsnAspGluLysAspGluGlnValAlaAsnLysG 34
18462 CAACCCAGGAAATGATGAAAGAGATGAAAGAGCAAGTCTGCTAATAAG 18511

34 lyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
18512 GGGAGCCCTTGGCCCTACCTTGAATGTTAGTGAATCTGTGCTCTAGA 18561

51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
18562 GGAACCGTAGCGGTTCCGGTTAGGAGGCCATCTGCGAGTATAGATG 18611

67 pAspIleMetHisArgLeuGluGluProGlnAlaArgMetArgGluGluA 84
18612 GGACATAATGATGCTTGGAGAGCCACAGGCAAGGATGAGAGAGGAGA 18661

84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
18662 ATATGGAAGGATTTGGGAGGAGTGAGACAGCTGATGGAAGAGCTGAGG 18711

101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 117
18712 GAAAAGCATGACATGATGAGTTTGGCGGAGTGCAGCTGATCCCTCTCA 18761

117 sHisAspHisAspGluPheCysLeuMetPro 128
18762 CCATGACCATCAGATGAGTTTGGCGGAGTGCAGCTGATCCCTCTCA 18795

seq_name: gb_pr:AF183416
seq_documentation_block:
LOCUS      AF183416 642 bp  mRNA  PRI  02-SEP-2000
```



```

DEFINITION Homo sapiens ovarian granulosa cell 13.0 kDa protein hGR74 homolog
mRNA, complete cds.
ACCESSION AF183416
VERSION AF183416.1 GI:9963770
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 642)
AUTHORS Peng, Y., Qian, B., Tu, Y., Xu, S., Han, Z., Fu, G. and Chen, Z.
TITLE A novel gene expressed in human adrenal gland
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 642)
AUTHORS Peng, Y., GU, W., Huang, C., Xu, S., Han, Z., Fu, G. and Chen, Z.
TITLE Direct Submission
JOURNAL Submitted (03-SEP-1999) Chinese National Human Genome Center at
Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong,
Shanghai 201203, P. R. China
FEATURES
source
1..642
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="adrenal gland"
3..380
/notes="HGR74-h protein"
/codon_start=1
/evidence="not_experimental"
/product="ovarian granulosa cell 13.0 kDa protein hGR74
homolog"
1..642
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="adrenal gland"
3..380
/notes="HGR74-h protein"
/codon_start=1
/evidence="not_experimental"
/product="ovarian granulosa cell 13.0 kDa protein hGR74
homolog"
CDS
1..642
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="adrenal gland"
3..380
/notes="HGR74-h protein"
/codon_start=1
/evidence="not_experimental"
/product="ovarian granulosa cell 13.0 kDa protein hGR74
homolog"
BASE COUNT 204 a 118 c 157 g 163 t
ORIGIN
1
MetGluseryLysGluArgAlaLeuAsnAsnLeuValGluAsnVa 17
|||||
3 ATGAGTCCAAAGAGAACGACGAGTAACAGCTCTCAGCATGGAATGC 52
|||||
17 lAsnGlnGluAsnAspGluLysAspGluGlnValAlaAsnLysG 34
|||||
53 CAACCAAGAAAT ..... GAAGAAAGGAGCAAGTTGCTAATAAAG 93
|||||
34 lYGluproLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
94 GGAGAGCCCTTGCCCTCTTGATGCTGGTGAATCTGTGTCCTAGA 143
|||||
51 GlyAsnArgArgPheArgValArgGlnProileLeuGlnTyrArgTr 67
|||||
144 GGAATCGTAGGCGGTTCGCGCTTAGCGAGCCCATCTCGCATATAGT 193
|||||
67 pAspIleMetHisArgLeuGluProGlnAlaArgMetArgGluAla 84
|||||
194 GGATATGATGATGATGCTTGAGAACACACAGGAGGATGAGAGAGA 243
|||||
84 snMetGluArgTrleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
|||||
244 ATATGAAAGGATTGGGAGGAGGTGAGACAGCTGATGGAAGACCTGAGG 293
|||||

```

```

101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProH1 117
|||||
294 GAAAGCAGCTTCAGTCATAGTCTCGGGCAGTCAGCACTGACCCCTCA 343
|||||
117 SHIsaspHisAspGluPheCysLeuMetPro 128
|||||
344 CCATGACCATCATGATGAGTTTTCCTTATGCC 377
|||||
seq_name: gb_pr:AF220189
seq_documentation_block: 828 bp mRNA PRI 04-MAY-2000
LOCUS AF220189
DEFINITION Homo sapiens uncharacterized hypothalamus protein HBEX2 mRNA,
complete cds.
ACCESSION AF220189
VERSION AF220189.1 GI:7689028
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 828)
AUTHORS Xiao, H., Song, H., Gao, G., Ren, S., Chen, Z. and Han, Z.
TITLE A novel gene expressed in human hypothalamus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 828)
AUTHORS Xiao, H., Song, H., Gao, G., Ren, S., Chen, Z. and Han, Z.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Chinese National Human Genome Center at
Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong,
Shanghai 201203, P. R. China
FEATURES
source
1..828
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="hypothalamus"
167..544
/codon_start=1
/product="uncharacterized hypothalamus protein HBEX2"
/protein_id="AAF6754.1"
/db_xref="GI:7689028"
/translation="MESKEKRAVNSLSMENAOEKEQVANKGEPLALPLDAGEYC
VPGNRRFRVROPILQYRWDMHRLGEPQARMREENMERIGEVRQMLEKLEKQLS
HSLRAVSTDPHHDHDECLMP"
BASE COUNT 256 a 170 c 221 g 181 t
ORIGIN
1
MetGluseryLysGluArgAlaLeuAsnAsnLeuValGluAsnVa 17
|||||
167 ATGAGTCCAAAGAGAACGACGAGTAACAGCTCTCAGCATGGAATGC 216
|||||
17 lAsnGlnGluAsnAspGluLysAspGluGlnValAlaAsnLysG 34
|||||
217 CAACCAAGAAAT ..... GAAGAAAGGAGCAAGTTGCTAATAAAG 257
|||||
34 lYGluproLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
258 GGAGAGCCCTTGCCCTCTTGATGCTGGTGAATCTGTGTCCTAGA 307
|||||
51 GlyAsnArgArgPheArgValArgGlnProileLeuGlnTyrArgTr 67
|||||
308 GAAATCGTAGGCGGTTCGCGCTTAGCGAGCCCATCTCGCATATAGT 357
|||||

```

alignment_scores:

Quality: 606.50 Length: 128
Ratio: 4.891 Gaps: 1
Percent Similarity: 96.875 Percent Identity: 89.062

alignment_block:

US-09-327-750D-31 x AF220189

Align seg 1/1 to: AF220189 from: 1 to: 828

```

1 MetGluseryLysGluArgAlaLeuAsnAsnLeuValGluAsnVa 17
|||||
167 ATGAGTCCAAAGAGAACGACGAGTAACAGCTCTCAGCATGGAATGC 216
|||||
17 lAsnGlnGluAsnAspGluLysAspGluGlnValAlaAsnLysG 34
|||||
217 CAACCAAGAAAT ..... GAAGAAAGGAGCAAGTTGCTAATAAAG 257
|||||
34 lYGluproLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
258 GGAGAGCCCTTGCCCTCTTGATGCTGGTGAATCTGTGTCCTAGA 307
|||||
51 GlyAsnArgArgPheArgValArgGlnProileLeuGlnTyrArgTr 67
|||||
308 GAAATCGTAGGCGGTTCGCGCTTAGCGAGCCCATCTCGCATATAGT 357
|||||

```


repeat_region 29603..30740
/note="RIGGER1 repeat: matches 587..1757 of consensus"
repeat_region 31027..31365
/note="RIGGER1 repeat: matches 1757..2105 of consensus"
repeat_region 31664..31691
/note="RIGGER1 repeat: matches 2105..2132 of consensus"
repeat_region 31809..32015
/note="RIGGER1 repeat: matches 2132..2246 of consensus"
repeat_region 32323..32389
/note="RIGGER1 repeat: matches 2246..2313 of consensus"
repeat_region 34353..34636
/note="MLT1-INTERNAL repeat: matches 607..889 of consensus"
repeat_region 35149..35376
/note="MLT1-INTERNAL repeat: matches 1001..1233 of consensus"
repeat_region 35523..35670
/note="MLT1C repeat: matches 8..159 of consensus"
BASE COUNT 9708 a 7094 c 6805 g 12107 t
ORIGIN

alignment_scores:
Quality: 606.50 Length: 128
Ratio: 4.891 Gaps: 1
Percent Similarity: 96.875 Percent Identity: 89.062
alignment_block:
US-09-327-750D-31 x H5198P4/rev ..
Align seg 1/1 to reverse of: H5198P4 from: 1 to: 35714

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
|||||
24358 ATGGAGTCCAAAGAGAACGAGCAGTAAACAGCTCTCAGCATGGAATGCG 24309
|||||
17 lasnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
|||||
24308 CAACCAAGAAAT.....GAAGAAAGAGGAGCAAGTTGCTAATAAG 24268
|||||
34 lyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
24267 GGGAGCCCTTGGCCCTCCCTTGGATGCTGGTGAATAGTGTGCGCTAGA 24218
|||||
51 GlyAsnArgArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
|||||
24217 GGAATATCTAGCGGTTCCGGTTAGGAGCCATCTCTGAGTATAGATG 24168
|||||
67 paspleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84
|||||
24167 GGATATGATGCTAGGCTTGGAGAACCCACAGGCAAGGATGAGAGAAGA 24118
|||||
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
|||||
24117 ATATGAAAGGATTGGGAGGAGGTGAGACAGCTGTGGAAGAGCTGAGG 24068
|||||
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 117
|||||
24067 GAAAGACAGTTGAGTCATAGTCTCGGGCAGTCAGCACTGACCCCTCA 24018
|||||
117 shIsAspHisHisAspGluPheCysLeuMetPro 128
|||||
24017 CCATGACCATCATGATGATGTTTGCCTTATGCC 23984

seq_name: gb_pr:AF237783
seq_documentation_block:
LOCUS AF237783 791 bp mRNA PRI 02-SEP-2000
DEFINITION Homo sapiens brain-expressed protein BEX1 (BEX1) mRNA, complete cds.
ACCESSION AF237783
VERSION AF237783.1 GI:9963898
KEYWORDS

SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 791)
AUTHORS Yang,Q.S., Ying,K., Xie,Y. and Mao,Y.M.
TITLE A Novel Human X-linked Brain Expressed Protein
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 791)
AUTHORS Mao,Y.M., Xie,Y., Yang,Q.S., Wu,H., Lin,S. and Ying,K.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2000) Genetic Research Institute, Fudan University, 220 Handan Rd., Shanghai 200433, P.R.China
FEATURES
Location/Qualifiers
1..791
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="between DXS990 and DX1059 by RH (Stanford G3)"
1..791
/gene="BEX1"
172..549
/gene="BEX1"
/codon_start=1
/product="brain-expressed protein BEX1"
/protein_id="AAG09752.1"
/db_xref="GI:9963898"
/translation="MESKEKLAVNSLMSNANOEKEOVANKGEPLALPLDAGEVC
VPGNRFRFRVQPIQYRNDMHRLGEPQARMREENMERIGESVRLMEKLEKQLS
HSLRVSTDPHHDHDEFCLMP"
BASE COUNT 214 a 172 c 218 g 187 t
ORIGIN

alignment_scores:
Quality: 592.50 Length: 128
Ratio: 4.857 Gaps: 1
Percent Similarity: 95.312 Percent Identity: 87.500
alignment_block:
US-09-327-750D-31 x AF237783 ..
Align seg 1/1 to: AF237783 from: 1 to: 791
1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
|||||
172 ATGGAGTCCAAAGAGAACTAGCAGTAAACAGCTCTCAGCATGGAATGCG 221
|||||
17 lasnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
|||||
222 CAACCAAGAAAT.....GAAGAAAGAGGAGCAAGTTGCTAATAAG 262
|||||
34 lyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
263 GGGAGCCCTTGGCCCTCCCTTGGATGCTGGTGAATAGTGTGCGCTAGA 312
|||||
51 GlyAsnArgArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
|||||
313 GGAAATCTAGCGGTTCCGGCTTAGGAGCCATCTCTGAGTATAGATG 362
|||||
67 paspleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84
|||||
363 GGATATGATGCTAGGCTTGGAGAACCCACAGGCAAGGATGAGAGAAGA 412
|||||
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
|||||
413 ATATGAAAGGATTGGGAGGAGGTGAGACAGCTGTGGAAGAGCTGAGG 462
|||||
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 117
|||||
463 GAAAGACAGTTGAGTCATAGTCTCGGGCAGTCAGCACTGACCCCTCA 512
|||||
117 shIsAspHisHisAspGluPheCysLeuMetPro 128

```

|||||
513 CCATGACATCATGATGAGTTTGGCTTATGCC 546

seq_name: gb_ro:AF097439
seq_documentation_block:
LOCUS AF097439 785 bp mRNA ROD 13-APR-1999
DEFINITION Mus musculus brain expressed X-linked protein 2 (Bex2) mRNA,
complete cds.
ACCESSION AF097439
VERSION AF097439.1 GI:4580591
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 785)
JOURNAL Brown, A.L. and Kay, G.F.
MEDLINE Bex1, a gene with increased expression in parthenogenetic embryos,
REMARK Hum. Mol. Genet. 8 (4), 611-619 (1999)
Erratum: [[published erratum appears in Hum Mol Genet 1999
May; 8(5):943]]
REFERENCE 2 (bases 1 to 785)
AUTHORS Brown, A.L. and Kay, G.F.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1998) Cancer Unit, Queensland Institute of
Medical Research, Herston Rd, Brisbane, Qld 4029, Australia
FEATURES
source
1. 785
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="dbEST:W48832"
/chromosome="X"
/map="near Pip"
/dev_stage="embryo; 15.5 dpc"
1. 785
/gene="Bex2"
139..528
/gene="Bex2"
/codon_start=1
/product="brain expressed X-linked protein 2"
/protein_id="AAD24430.1"
/db_xref="GI:4580592"
/translation="MESKVEQGVKNLMNENDHOKKEEKPDASKRDPTVALPFEA
GDYVPRGRRRFRVROPVHVYRWDLMDLQVRGEPQGRMRREENVQFGDDVRLMEKLRER
ROLSHSLRAVSTDPHHDHDFCLMP"
BASE COUNT 212 a 163 c 235 g 175 t
ORIGIN

alignment_scores:
Quality: 502.50 Length: 129
Ratio: 4.153 Gaps: 1
Percent Similarity: 93.798 Percent Identity: 71.318

alignment_block:
US-09-327-750d-31 x AF097439
Align seg 1/1 to: AF097439 from: 1 to: 785
1 MetGlusLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
|||||
139 ATGAGTCCAAAGTGAACAGCGCTGAAATCTCAACATGGAGATGA 188
|||||
17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
|||||
189 CCATCAGGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 238
|||||
34 lYcGluProLeu...AlaLeuProLeuAsnValSerGluTyrCysValPro 49
|||||

```

```

239 GGGATCCGATTGTGGCCCTCCCTTTTCGAAGCTGGAGACTACTACGTGCCT 288
50 ArgGlyAsnArgArgArgPheArgValArgGlnProIleLeuGlnTyrAr 66
|||||
289 AGAGGAGTCCGAGCGGTTCCGGGTCGCGAGCCCTCGTGCCTACAG 338
66 gTTPAspIleMetHisArgLeuGluGluProGlnAlaArgMetArgGluG 83
|||||
339 ATGGACCTGATGCTAGGTGGGAGCCGCCAGGAGGAGGAGGAGGAGG 388
83 luAsnMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeu 99
|||||
389 AGAACGTACAGAGGTTTGGGATGATGTGAGACAGCTCATGAGAAGCTG 438
100 ArgGluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
|||||
439 AGGAAGGAGCTGAGCCACAGCTCGCGGGGTAGCACTGACCCGCC 488
116 OHSHISASPHISHisAspGluPheCysLeuMetPro 128
|||||
489 TCATCATGACCACTGATGAGTTTGCCTTATGCC 525
seq_name: gb_ro:AF097437
seq_documentation_block:
LOCUS AF097437 2269 bp DNA ROD 05-AUG-1999
DEFINITION Mus musculus Bex1 protein (Bex1) gene, complete cds.
ACCESSION AF097437
VERSION AF097437.1 GI:5702155
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 2269)
JOURNAL Brown, A.L. and Kay, G.F.
MEDLINE Bex1, a gene with increased expression in parthenogenetic embryos,
REMARK Hum. Mol. Genet. 8 (4), 611-619 (1999)
99172070
10072429
REFERENCE 2 (bases 1 to 2269)
AUTHORS Brown, A.L. and Kay, G.F.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1998) Cancer Unit, Queensland Institute of
Medical Research, Herston Rd, Brisbane, Qld 4030, Australia
FEATURES
source
1. 2269
/organism="Mus musculus"
/strain="129SVJ"
/db_xref="taxon:10090"
/chromosome="X"
/map="near Pip"
/tissue_type="liver"
/join(409..509,866..942,1263..1873)
/gene="Bex1"
/product="Bex1 protein"
409..1873
/gene="Bex1"
/note="expressed in brain; X-linked"
1268..1654
/gene="Bex1"
/codon_start=1
/product="Bex1 protein"
/protein_id="AAD47168.1"
/db_xref="GI:5702156"
/translation="MESKDGQGVKNLMNENDHOKKEEKPDQTTIRREPAVALTSEAG
KNCAPRGRRRFRVROPVIAHVRWDLMDLQVRGEPQGRMRREENVQFGDDVRLMEKLRER
QLSHSLRAVSTDPHHDHDFCLMP"
BASE COUNT 563 a 554 c 705 g 447 t
ORIGIN

```


BC003254	2901 bp	mus mus	20-FEB-2001
Mus musculus, Similar to dentatorubral pallidolysian atrophy, clone MGC:5758, mRNA, complete cds.			

DCO05245.1 GI.13090500
 MGC.
 house mouse.
 Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
 1 (bases 1 to 2901)

Strausberg, R.
Direct Submission
Submitted (20-FEB-2001) National Institutes
Genetic Collection (MGC), Cancer Genomics Office,
Institute, 31 Center Drive, Room 11A03, Bethesda,
USA
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: Robert Strausberg, Ph.D.

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

Tel: (301) 496-1330
 Email: Robert_Strausberg@nih.gov

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center

Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC plate: 9 Row: 0 Column: 6.

Location/Qualifiers
1. .2901

```
/clone= M03:5/58 IMAGE:3300522"
/tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating
ductal carcinoma. 5 month old virgin mouse."
```

```

/clone_lib="NCI_CGAP_Mam6"
/lab_host="DH10B"
/vec="Vector"
/notes="Vector"

```

```

/product="Similar to dentatorubral pallidoluysian atrophy"
/protein_id="AAH03254.1"
/db_xref="GI:13096907"

```

```
/translation="MVPQELLPLPGLPKPPIHVHTPLHPRAPKEHPTPSHQSLQSP  
PQLPPLSSPPAAPTGPPLTATQIKQPAEEYEPEESPVPARSPPPPKVVDPVS  
HASASARENKHI.DRGENSECARSTVEVDIECSKIAYKDANIYGVWUBBBAEDDAE
```

REREREKEREREKELERSVKLAQGRAPVECPSLGPPHRRPPEPGSAVATVPP
YLGPDTPALRTLSEYARPHVMSPGNRNHPFYVPLGAVDPGLLGYNPVALYSSDPAARE

KEKEAKEKOLDKRLKPGFEVAPSELEPLHGVPGPGLDPPPHGGLALQPGPPGLHPFP
FHPSLGLERERLALAAGPALRPDMSYAERLAAERQHAERVAALGNDPLARLQMLNVT
PHHHQSHIHSIHLHQQDATHAASASVHPLIDPLASGSHLTRIPYPAGTLPNPLLPB

PLHENEVLRHQLFAPYPYRIDIPASLSAPMSAAHQIQAMHAQSAELQRLALEQQQWLHAA
HPLHSVPLPAQEDYYSHLKKESDKPL"
623 a 958 c 791 g 529 t

cores:
Quality: 388.00 Length: 119

Ratio:	3.660	Gaps:	2
Similarity:	89.076	Percent Identity:	63.866

Block: 50D-31 x BC003254 ..

US-09-327-750D-31 x BC003254

Align seg 1/1 to: BC003254 from: 1 to: 2901

```

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
223 ATGGAGTCCAAAGATCAA...GGCGTGAATAATCTCAACATGCGAATCA 269
17 IAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
270 CCATCAGAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 319
34 IyGluPro...LeuAlaLeuProLeuAsnValSerGluTyrCysValPro 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
320 GGGAGCCAGCTGTGGCCCTGACCTCCAGGCTGGCAAAACTGTGCACCT 369
50 ArgGlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrAr 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
370 AGAGGAGGTCTGAGCGGTTCCGGGTTCCGGCAGCCCATCGCTACATAG 419
66 gTrpAspIleMetHisArgLeuGluGluProGlnAlaArgMetArgGluG 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
420 ATGGACCTGATGTCAGAGGTTGGGGAGGCGCCAGGAGGAGGAGGAGG 469
83 IuAsnMetGluArgIleGluGluValArgGlnLeuMetGluLysLeu 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
470 AGAACGTACAGAGGTTGGGGTCTATCTGAGCAGCTCATGGAGAGCTG 519
100 ArgGluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
520 AGGAAAGGCGAGCTGAGCCAGCTCGGGCGGTTAGCTAGCTAGCCGCC 569
116 oHishis 118
|||||
570 TCATCAT 576

```

seq_name: gb_sts:G24641

seq_documentation_block: 504 bp DNA STS 31-MAY-1996
 LOCUS G24641
 DEFINITION human STS WI-11354, sequence tagged site.
 ACCESSION G24641
 VERSION G24641.1 GI:1344967
 KEYWORDS STS; STS sequence; primer; sequence tagged site.
 SOURCE human STSs derived from sequences in dbEST and the Unigene collection.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 504)
 AUTHORS Hudson,T.
 TITLE Whitehead Institute/MIT Center for Genome Research; Physically
 Mapped STSS
 JOURNAL Unpublished (1995)

Contact: Thomas Hudson
 Whitehead Institute/MIT Center for Genome Research
 Whitehead Institute for Biomedical Research
 9 Cambridge Center, Cambridge MA 02142 USA
 Tel: 617 252 1900
 Fax: 617 252 1902
 Email: thudson@genome.wi.mit.edu

Primer A: TTTTTCGCTTTTAAAGGTGC
 Primer B: TTACGTGGGTCTCTTATACCA
 STS size: 150
 PCR Profile:

Presoak:
 Denaturation:
 Annealing: 56 degrees C
 Polymerization:
 PCR Cycles: 35
 Thermal Cycler:
 Protocol:
 Template: 10 ng

Primer: each 5 pM
 dNTPs: each 4 nM
 Taq Polymerase: 0.025 units/ul
 Total Vol: 20 ul

Buffer:
 MgCl2: 1.5 mM
 KCl: 50 mM
 Tris-HCL: 10 mM
 pH: 9.3

Derived from dbEST (genbank accession T86927).
 Location/Qualifiers
 1..504

FEATURES
 source
 1..504
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

STS
 primer_bind
 1..150
 1..22

primer_bind
 complement(129..150) 144 t 7 others
 BASE COUNT 134 a 118 c 101 g 144 t 7 others
 ORIGIN

alignment_scores:
 Quality: 339.00 Length: 85
 Ratio: 4.644 Gaps: 3
 Percent Similarity: 85.882 Percent Identity: 83.529

alignment_block:
 US-09-327-750D-31 x G24641/rev

Align seg 1/1 to reverse of: G24641 from: 1 to: 504

```

47 CysValProArgGly.AsnArgArgArgPhe.ArgValArgGlnProIle 62
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
503 TGTCCCCCTANAGGAANCCGTAGCGGTTCCCCCGTTAGGCANCCCCAC 454
63 LeuGlnTyrArgTrpAspIleMetHisArgLeuGlu.ProGlnAlaA 79
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
453 CCGCAGTANAGATGGGATATGATGATAGGTGGAGAACCCACAGCAA 404
79 IqMetArgGluGluAsnMetGluArgIleGlyGluGluValArgGlnLeu 95
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
403 GGATGAGAGAAAGAGATATGGAAGGATTGGGAGAGGTGAGACAGCTG 354
96 MetGluLysLeuArgGluLysGlnLeuSerHisSerLeuArgAlaVal 112
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
353 ATGGAAAAGCTGAGGGAAGACAGTTGAGTCANAGTCTGGCGGCGAGTCAG 304
112 rThrAspProHisHisAspHisHisAspGluPheCysLeuMetPro 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
303 CACTGACCCCCCTCACCATGACCATCATGATGAGTTTGTGTTATGCC 255

```

OM of: US-09-327-750D-32 to: EST:* out_format : pfs

Date: Mar 11, 2002 2:16 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet-p2n.model -DEV=xlh
-Q=cn22.1/USPQ-spool/US09327750/runat_11032002_101153_20308/app_query.fasta_1.1472
-DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.050 -LOOPECL=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blotsum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODEL=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZ=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09327750 -CGN1_15654
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-327-750D-32

Query length: 125

Database: EST*

Database sequences: 11351937

Database length: 1077921985

Search time (sec): 4085.940000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_est1:BG655001	+	653.00	1221.08	547	1 BG655001 1b44f06.y1 HR85 islet
gb_est2:BG472167	+	653.00	1219.74	634	1 BG472167 602513894F1 NIH_MGC_16
gb_est3:BF971303	+	653.00	1219.40	9.3e-59	1 BF971303 602273150F1 NIH_MGC_84
gb_est4:BG714823	+	653.00	1219.18	9.3e-59	1 BG714823 602677146F1 NIH_MGC_98
gb_est5:AV702285	+	653.00	1218.97	9.8e-59	1 AV702285 AV702285 ADB Homo sapi
gb_est6:AV705808	+	653.00	1218.96	9.9e-59	1 AV705808 AV705808 ADB Homo sapi
gb_est7:AL526209	+	653.00	1218.91	9.9e-59	1 AL526209 AL526209 LTI_NFL003.NE
gb_est8:BG820179	+	653.00	1218.85	1.0e-58	1 BG820179 602782222F1 NCI_CGAP_E
gb_est9:BG790774	+	653.00	1217.82	1.1e-58	1 BG790774 601581956F1 NIH_MGC_7
gb_est10:AL526247	+	653.00	1217.70	1.2e-58	1 AL526247 AL526247 LTI_NFL003.NE
gb_est11:AV702643	+	653.00	1217.43	2.0e-58	1 AV702643 AV702643 ADB Homo sapi
gb_est12:AL929703	+	646.00	1207.46	4.3e-58	1 AL929703 au63f04.y1 Schneider f
gb_est13:BG715659	+	646.00	1205.26	5.7e-58	1 BG715659 602676924F1 NIH_MGC_98
gb_est14:BG709325	+	637.00	1187.25	5.8e-57	1 BG709325 602673345F1 NIH_MGC_98
gb_est15:AW160832	+	631.00	1180.01	1.5e-56	1 AW160832 au76b12.y1 Schneider f
gb_est16:BG346478	+	630.00	1176.22	2.4e-56	1 BG346478 602020342F1 NCI_CGAP_E
gb_est17:AV727777	+	627.00	1171.87	4.1e-56	1 AV727777 AV727777 HTC Homo sapi
gb_est18:AW553922	+	627.00	1171.28	4.5e-56	1 AW553922 EST365887 MAGE resequ
gb_est19:BG701861	+	627.00	1168.97	6.0e-56	1 BG701861 602883214F1 NIH_MGC_95
gb_est20:AV728527	+	626.00	1164.93	1.0e-55	1 AV728527 AV728527 HTC Homo sapi
gb_est21:BF966650	+	625.00	1165.09	9.9e-56	1 BF966650 602286982F1 NIH_MGC_95
gb_est22:BG714687	+	607.00	1132.41	6.5e-54	1 BG714687 602677046F1 NIH_MGC_98
gb_est23:BG314909	+	602.50	1125.45	1.6e-53	1 BG314909 601140074F1 NIH_MGC_9
gb_est24:AL520931	+	602.50	1123.07	2.0e-53	1 AL520931 AL520931 LTI_NFL004.NE
gb_est25:BG889313	+	602.50	1123.74	2.2e-53	1 BG889313 601513201F1 NIH_MGC_71
gb_est26:BG709503	+	602.50	1122.63	2.3e-53	1 BG709503 602674781F1 NIH_MGC_98
gb_est27:BG703798	+	602.50	1120.35	2.4e-53	1 BG703798 602672810F1 NIH_MGC_98
gb_est28:BG703877	+	602.50	1120.34	3.1e-53	1 BG703877 601462119F1 NIH_MGC_98
gb_est29:BG285825	+	600.50	1117.46	4.4e-53	1 BG285825 602380917F1 NIH_MGC_95
gb_est30:BG036775	+	600.50	1118.04	4.1e-53	1 BG036775 602286982F1 NIH_MGC_95
gb_est31:CI15948	+	598.00	1118.78	3.7e-53	1 CI15948 C15948 Clontech human ad
gb_est32:BF348593	+	598.00	1113.14	7.7e-53	1 BF348593 602019889F1 NCI_CGAP_E
gb_est33:AW024587	+	597.50	1115.16	6.0e-53	1 AW024587 wu77g08.x1 NCI_CGAP_K4
gb_est34:BG714974	+	595.50	1110.47	1.1e-52	1 BG714974 602677263F1 NIH_MGC_98
gb_est35:BF967675	+	595.50	1109.67	1.2e-52	1 BF967675 602287783F1 NIH_MGC_98
gb_est36:BG707734	+	595.50	1109.40	1.2e-52	1 BG707734 602671255F1 NIH_MGC_98
gb_est37:AA160657	+	594.00	1108.81	1.3e-52	1 AA160657 zq49f12.r1 Stratagene
gb_est38:W48778	+	592.00	1107.91	1.5e-52	1 W48778 zc44c03.r1 Soares_senes
gb_est39:BG705843	+	591.50	1101.93	3.3e-52	1 BG705843 602669329F1 NIH_MGC_98
gb_est40:AI291270	+	589.50	1102.39	3.1e-52	1 AI291270 qml6e06.x1 NCI_CGAP_LU

gb_est1:AI291126 + 587.50 1096.19 6.8e-52 669 1 AI291126 qml5f02.x1 NCI_CGAP
gb_est2:BF237433 + 587.50 1095.63 7.3e-52 711 1 BF237433 601842108F1 NIH_MGC
gb_est3:BE266012 + 586.50 1095.14 7.8e-52 610 1 BE266012 601194731F1 NIH_MGC
gb_est4:BE733217 + 584.50 1087.88 2.0e-51 897 1 BE733217 601568432F1 NIH_MGC
gb_est5:BF967457 + 581.50 1083.65 3.4e-51 769 1 BF967457 602287383F1 NIH_MGC

seq_name: gb_est2:BG655001

seq_documentation_block: 547 bp mRNA EST 05-JUL-2001
LOCUS BG655001
DEFINITION 1b44f06.y1 HR85 islet Homo sapiens cDNA 5' similar to FR:QWRT28
QWRT28 BRAIN EXPRESSED X-LINKED PROTEIN 2.; mRNA sequence.
ACCESSION BG655001
VERSION BG655001.1 GI:13792410
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 547)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
M., Gibbons,M., McGann,R., Cole,R., Tsagarishvili,R., Williams,T.,
Jackson,Y. and Bowers,Y.
ENDOCRINE PANCREAS CONSORTIUM
UNPUBLISHED (2000)
Other ESTs: 1b44f06.x1

TITLE Endocrine Pancreas Consortium

JOURNAL

COMMENT

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@biohph.harvard.edu

Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:

Washington University Genome Sequencing Center For information on

obtaining a clone please contact: Dr. Hiroshi Inoue

(hinoue@wustl.edu)

Seq primer: -40RP from Gibco

High quality sequence stop: 447.

FEATURES

Location/Qualifiers

1..547

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="HR85 islet"

/tissue_type="Purified pancreatic islet"

/lab_host="DH10B"

/note=Organ: Pancreas; Vector: pBluescript SK(-); Site_1:

NotI; Site_2: XhoI; cDNA made by oligo-dr priming.

Size-selected on agarose gel. Average insert size -1kb. 5'

XhoI site was destroyed after directional cloning.

Amplified once. Contact information: Hiroshi Inoue, MD,

Metabolism Div. (Alan Permutt Lab), Washington University

School of Medicine, Box 8127, 660 South Euclid Ave., St.

Louis, MO 63110, E-mail: hinoue@ingate.wustl.edu, Tel:

314-362-1916, Fax: 314-747-2692."

BASE COUNT 152 a 120 c 174 g 101 t

ORIGIN

alignment_scores:

Quality: 653.00 Length: 125

Ratio: 5.266 Gaps: 0

Percent Similarity: 99.200 Percent Identity: 98.400

alignment_block:

US-09-327-750D-32 x BG655001

Align seg 1/1 to: BG655001 from: 1 to: 547

(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH/MGC Library."

BASE COUNT
ORIGIN

alignment_scores:		
Quality:	653.00	Length: 125
Ratio:	5.266	Gaps: 0
Percent Similarity:	99.200	Percent Identity: 98.400

alignment_block:

US-09-327-750D-32 x BG472167

Align seq 1/1 to: BG472167 from: 1 to: 634

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
 |||||
 30 ATGGAGTCCAAAGAGAAACGAGCAGTAAACAGTCTCAGCATGGAAATGC 79

seq_documentation_block:
LOCUS BG472167 634 bp mRNA EST 21-MAR-2001
DEFINITION 602513894F1 NIH_MGC_16 Homo sapiens CDNA IMAGE:4645775 5',
mRNA sequence.

Accession	Version	GI
BG472167	BG472167.1	GI:13404541

SOURCE	ORGANISM	Human:
	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	

REFERENCE 1 (Dates 1 to 634)
 AUTHORS Nih-MGC <http://mgc.ncl.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.

FEATURES

```

location/qualifiers
1. .634
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:4645775"
/clone_lib="NIH_MGC_16"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pOT87; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGACGACG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit

```

http://image.llnl.gov
Plate: LLAM10003 row: d column: 14
High quality sequence stop: 658.
Location/Qualifiers

FEATURES

source
1. .658
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4361053"
/clone_lib="NIH_MGC_84"
/tissue_type="adrenal cortex carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site: 1:
primed. Average insert size 1.229 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 205 a 124 c 164 g 165 t
ORIGIN

alignment_scores:
Quality: 653.00 Length: 125
Ratio: 5.266 Gaps: 0
Percent Similarity: 99.200 Percent Identity: 98.400

alignment_block:

US-09-327-750D-32 x BF971303

Align seg 1/1 to: BF971303 from: 1 to: 658

```

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||
24 ATGGAGTCCAAAGAGAAAGAGCAGGTAAACAGCTCTCAGCATGGAATGC 73
|||||
17 aAsnGlnGluAsnGluLysGluGlnValAlaAsnLysGlyGluProL 34
|||||
74 CAACCAAGAAATGAAGAAAGAGCAAGTTGCTAATAAAGGGAGCCCT 123
|||||
34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50
|||||
124 TGGCCCTCCCTTTGGAGTCTGGTGAATAGTGTGCTAGAGAAATCGT 173
|||||
51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIle 67
|||||
174 AGGCGGTTCGCGTTAGCGAGCCATCCTGCAGTATAGATGGGATATGAT 223
|||||
67 rHisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGlu 84
|||||
224 GCATAGGCTTGGAGAACACAGCAGGAGGATGAGAGAGATATGGAAA 273
|||||
84 rGileGlyGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100
|||||
274 GGATGGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGGAAAAGCAG 323
|||||
101 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspHi 117
|||||
324 TTGAGTCATAGCTGCGGGCAGTCAGCACTGACCCCTCACCATGACCA 373
|||||
117 shiAspGluPheCysLeuMetPro 125
|||||
374 TCATGATGAGTTTGCCTTATGCC 398

```

seq_name: gb_est2:BG714823

seq_documentation_block:

LOCUS BG714823 674 bp mRNA EST 08-MAY-2001
DEFINITION 602677146F1 NIH_MGC_96 Homo sapiens CDNA clone IMAGE:4799747 5',
mRNA sequence.
ACCESSION BG714823
VERSION BG714823.1 GI:13993754
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE

AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 674)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10689 row: k column: 12
High quality sequence stop: 672.

FEATURES

source
1. .674

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4799747"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"

/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site: 1: BamHI; Site: 2: SalI-xhoI (gtcag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.3 kb and
normalized to 10^5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 175 a 156 c 201 g 142 t
ORIGIN

alignment_scores:

Quality: 653.00 Length: 125
Ratio: 5.266 Gaps: 0
Percent Similarity: 99.200 Percent Identity: 98.400

alignment_block:

US-09-327-750D-32 x BG714823

Align seg 1/1 to: BG714823 from: 1 to: 674

```

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||
173 ATGGAGTCCAAAGAGAAAGAGCAGGTAAACAGCTCTCAGCATGGAATGC 222
|||||
17 aAsnGlnGluAsnGluLysGluGlnValAlaAsnLysGlyGluProL 34
|||||
223 CAACCAAGAAATGAAGAAAGAGCAAGTTGCTAATAAAGGGAGCCCT 272
|||||
34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50
|||||
273 TGGCCCTCCCTTTGGAGTCTGGTGAATAGTGTGCTAGAGAAATCGT 322
|||||
51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIle 67
|||||
323 AGGCGGTTCGCGTTAGCGAGCCATCCTGCAGTATAGATGGGATATGAT 372
|||||
67 rHisArgLeuGlyGluValArgGlnLeuMetGluLysLeuArgGluLysGln 84
|||||
373 GCATAGGCTTGGAGAACACAGCAGGAGGATGAGAGAGATATGGAAA 422
|||||
84 rGileGlyGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100
|||||
423 GGATGGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGGAAAAGCAG 472

```

101 LeuSerHisSerLeuArgAlaValSerThrAspProProHisHisAspHI 117
 473 TTGAGTCATAGTCTGGGGCAGTCAGCACTGACCCCTCACCATGACCA 522
 117 sHisAspGluPheCysLeuMetPro 125
 523 TCATGATGAGTTTGGCTTATGCC 547

seq_name: gb_est1:AV702285

seq_documentation_block:
 LOCUS AV702285 690 bp mRNA EST 08-OCT-2000
 DEFINITION AV702285 ADB Homo sapiens cDNA clone ADBXE07 5', mRNA sequence.
 ACCESSION AV702285
 VERSION AV702285.1 GI:10718615
 EST.
 KEYWORDS human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 690)
 AUTHORS Peng, Y., Song, H., Huang, Q., Huang, C., Gu, Y., Yang, Y., Gao, G., Xiao, H., Xu, X., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R., Chen, J., Chen, Z. and Han, Z.
 TITLE Homo sapiens cDNA ADB clones
 JOURNAL Unpublished (2000)
 COMMENT Contact: Zequang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.

FEATURES
 source
 1..690
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="ADBBXE07"
 /clone_lib="ADB"
 /tissue_type="Adrenal gland"
 /dev_stage="Adult"
 /lab_host="SOLR"
 /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 210 a 132 c 180 g 166 t 2 others
 ORIGIN
 alignment_scores:
 Quality: 653.00 Length: 125
 Ratio: 5.266 Gaps: 0
 Percent Similarity: 99.200 Percent Identity: 98.400
 alignment_block:
 US-09-327-750D-32 x AV702285

Align seg 1/1 to: AV702285 from: 1 to: 690

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
 60 ATGGAGTCCAAAGAAAGAACGAGCAGTAACAGTCTCAGCATGGAAATGC 109
 17 aAsnGlnGluAsnGluLysGluGlnValAlaAsnLysGlyGluProL 34
 110 CAACCAAGAAATCAAGAAAGAGCAAGTTCCTAATAAGGGGAGCCCT 159
 34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50
 160 TGGGCCCTCCCTTTGGATGGCTGGTGAATCTGTGCTTACAGGAATCGT 209
 51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 67

210 AGGCGGTTCCGCTTAGGAGCCCATCCCTGACGTATAGTGGCATATGAT 259
 67 tHisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 84
 260 GCATAGCTTGGAGAACACAGGAGGATGAGAGAAGAGATATGGAAA 309
 84 rGileGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100
 310 GGATTTGGGAGGAGGTGAGACCTGATGAAAGCTGAGGAAAGCAG 359
 101 LeuSerHisSerLeuArgAlaValSerThrAspProProHisHisAspHI 117
 360 TTGAGTCATAGTCTGGGGCAGTCAGCACTGACCCCTCACCATGACCA 409
 117 sHisAspGluPheCysLeuMetPro 125
 410 TCATGATGAGTTTGGCTTATGCC 434

seq_name: gb_est1:AV705808

seq_documentation_block:
 LOCUS AV705808 691 bp mRNA EST 09-OCT-2000
 DEFINITION AV705808 ADB Homo sapiens cDNA clone ADBCRH03 5', mRNA sequence.
 ACCESSION AV705808
 VERSION AV705808.1 GI:10723103
 EST.
 KEYWORDS human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 691)
 AUTHORS Peng, Y., Song, H., Huang, Q., Huang, C., Gu, Y., Yang, Y., Gao, G., Xiao, H., Xu, X., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R., Chen, J., Chen, Z. and Han, Z.
 TITLE Homo sapiens cDNA ADB clones
 JOURNAL Unpublished (2000)
 COMMENT Contact: Zequang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.

FEATURES
 source
 1..691
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="ADBCRH03"
 /clone_lib="ADB"
 /tissue_type="Adrenal gland"
 /dev_stage="Adult"
 /lab_host="SOLR"
 /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 211 a 133 c 181 g 162 t 4 others
 ORIGIN
 alignment_scores:
 Quality: 653.00 Length: 125
 Ratio: 5.266 Gaps: 0
 Percent Similarity: 99.200 Percent Identity: 98.400
 alignment_block:
 US-09-327-750D-32 x AV705808

Align seg 1/1 to: AV705808 from: 1 to: 691

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
 117

```

60 ATGGAGTCCAAAGAGAAACAGCAGTAAACAGTCTCAGCATGGAAATGC 109
17 aasnglncluasngluGluLysGluGlnValAlaAsnLysGlyGluProL 34
|||||
110 CAACCAAGAAATGAAGAAAGAGGAGCAAGTGTCTAATAAAGGGAGCCCT 159
|||||
34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50
|||||
160 TGGCCCTCCCTTGGATGCTGGTGAATCTGTGTGCTAGAGGAAATCGT 209
|||||
51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTyrAspIleMe 67
|||||
210 AGGCGGTTCCTGGTGGAGCCATCCCTGCAGTATAGATGGGATATGAT 259
|||||
67 thisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 84
|||||
260 GCATAGGCTTGGAGAACACAGCAAGGATGAGAGAGAGAAATATGGAAA 309
|||||
84 rgileGlyGluGluValArgGluLeuMetGluLysLeuArgGluLysGln 100
|||||
310 GGATTTGGGAGGAGGTGAGACACTGATGGAAGAAAGCTGAGGAAAGAG 359
|||||
101 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspHi 117
|||||
360 TTGAGTCATAGTCTGGCGGAGTGCAGCACTGACCCCTCACCATGACCA 409
|||||
117 shiAspGluPheCysLeuMetPro 125
|||||
410 TCATGATGAGTGTTCCTTATGCC 434

```

seq_name: gb_est1:AL526209

seq_documentation_block: 695 bp mRNA EST 13-FEB-2001
LOCUS AL526209 LTI_NFL003_NBC3 Homo sapiens cDNA clone CSDDC016FO01 3
DEFINITION prime, mRNA sequence.

ACCESSION AL526209
VERSION AL526209.1 GI:12789702
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr.

Location/Qualifiers

1. .695
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSDDC016FO01"
/clone_lib="LTI_NFL003_NBC3"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact: Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax: (1) 301 610
8371 Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com"

FEATURES

source

alignment_scores:

Quality: 653.00 Length: 125
Ratio: 5.266 Gaps: 0
Percent Similarity: 99.200 Percent Identity: 98.400

alignment_block:

US-09-327-750D-32 x AL526209/rev

Align seg 1/1 to reverse of: AL526209 from: 1 to: 695

```

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||
546 ATGGAGTCCAAAGAGAAACAGCAGTAAACAGTCTCAGCATGGAAATGC 497
|||||
17 aasnglncluasngluGluLysGluGlnValAlaAsnLysGlyGluProL 34
|||||
496 CAACCAAGAAATGAAGAAAGAGGAGCAAGTGTCTAATAAAGGGAGCCCT 447
|||||
34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50
|||||
446 TGGCCCTCCCTTGGATGCTGGTGAATCTGTGTGCTAGAGGAAATCGT 397
|||||
51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTyrAspIleMe 67
|||||
396 AGGCGGTTCCTGGTGGAGCCATCCCTGCAGTATAGATGGGATATGAT 347
|||||
67 thisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 84
|||||
346 GCATAGGCTTGGAGAACACAGCAAGGATGAGAGAGAGAAATATGGAAA 297
|||||
84 rgileGlyGluGluValArgGluLeuMetGluLysLeuArgGluLysGln 100
|||||
296 GGATTTGGGAGGAGGTGAGACACTGATGGAAGAAAGCTGAGGAAAGAG 247
|||||
101 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspHi 117
|||||
246 TTGAGTCATAGTCTGGCGGAGTGCAGCACTGACCCCTCACCATGACCA 197
|||||
117 shiAspGluPheCysLeuMetPro 125
|||||
196 TCATGATGAGTGTTCCTTATGCC 172

```

seq_name: gb_est2:BG820179

seq_documentation_block:

LOCUS BG820179 699 bp mRNA EST 22-MAY-2001
DEFINITION 602782222F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4933012
5', mRNA sequence.

ACCESSION BG820179

VERSION BG820179.1 GI:14167766

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

1 (bases 1 to 699)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)

AUTHORS

Unpublished (1999)

JOURNAL

Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov

COMMENT

Tissue Procurement: David N. Louis, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLAM10859 row: 1 column: 05

High quality sequence stop: 695.

FEATURES

source

Location/Qualifiers

1. .699

BASE COUNT

ORIGIN

158 a 198 c 154 g 183 t 2 others

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4933012"
 /clone_lib="NCLCGAP_Brn67"
 /tissue_type="anaplastic oligodendroglioma with lp/19q loss"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: brain; Vector: pCMV-SpOrf6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.3 kb. Constructed by Life Technologies. Note: this is a NCLCGAP Library."
 198 a 138 c 191 g 172 t

BASE COUNT
 ORIGIN

alignment_scores:
 Quality: 653.00 Length: 125
 Ratio: 5.266 Gaps: 0
 Percent Similarity: 99.200 Percent Identity: 98.400

alignment_block:

US-09-327-750D-32 x BG820179

Align seg 1/1 to: BG820179 from: 1 to: 699

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
 |||||

94 ATGGAGTCCAAAGAGAAACGACGATTAACAGTCTCAGCATTGGAATGC 143
 |||||

17 aasnGlnGluAsnGluGluGlnValAlaAsnLysGlyGluProL 34
 |||||

144 CAACCAAGAAATGAAGAAAGGACGACGTTGCTAATAAGGGAGCCCT 193
 |||||

34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50
 |||||

194 TGGCCCTCCCTTTGGATGCTGGGAAATCTGTGCTCCTAGAGGAATCGT 243
 |||||

51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTyrAspIle 67
 |||||

244 AGCGGGTTCGGCTTAGGACGCCATCTGCAGTATAGATGGATATGAT 293
 |||||

67 tHisArgLeuGlyCysProGlnAlaArgMetArgGluGluAsnMetGluA 84
 |||||

294 GCATAGGCTTGGAGAACCAAGCAAGGATGAGAGAGAGATATGGAAA 343
 |||||

84 rgileGlyGluGluValArgGlnLeuMetGluLysLeuArgGlyGln 100
 |||||

344 GGATTGGGAGGAGGTGAGACAGCTGATGGAAGCTGAGGGAAGCAG 393
 |||||

101 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspHi 117
 |||||

394 TTGAGTCATAGTCATGCGGGGAGTCAGCATCTGACCCCTCACCATGACCA 443
 |||||

117 sHisAspGluPheCysLeuMetPro 125
 |||||

444 TCATGATGAGTTTGGCTTATGCC 468
 |||||

seq_name: gb_estl:BE790774

seq_documentation_block:

LOCUS BE790774 783 bp mRNA EST 20-SEP-2000
 DEFINITION 601581956F1 NIH_MGC-7 Homo sapiens cDNA clone IMAGE:3936613 5', mRNA sequence.

ACCESSION BE790774

VERSION BE790774.1 GI:10211972

KEYWORDS EST

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 783)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
 COMMENT

Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue procurement: DCTD/BTP
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LLCM780 row: 0 column: 14
 High quality sequence start: 3
 High quality sequence stop: 780

FEATURES

Location/Qualifiers

1..783

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3936613"
 /clone_lib="NIH_MGC-7"
 /tissue_type="small cell carcinoma"
 /cell_line="MGC3"

/lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
 214 a 169 c 219 g 181 t

BASE COUNT
 ORIGIN

alignment_scores:

Quality: 653.00 Length: 125
 Ratio: 5.266 Gaps: 0
 Percent Similarity: 99.200 Percent Identity: 98.400

alignment_block:

US-09-327-750D-32 x BE790774

Align seg 1/1 to: BE790774 from: 1 to: 783

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
 |||||

170 ATGGAGTCCAAAGAGAAACGACGATTAACAGTCTCAGCATGGAAATGC 219
 |||||

17 aasnGlnGluAsnGluGluGlnValAlaAsnLysGlyGluProL 34
 |||||

220 CAACCAAGAAATGAAGAAAGGACGACGTTGCTAATAAGGGAGCCCT 269
 |||||

34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50
 |||||

270 TGGCCCTCCCTTTGGATGCTGGGAAATCTGTGCTCCTAGAGGAATCGT 319
 |||||

51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTyrAspIle 67
 |||||

320 AGCGGGTTCGGCTTAGGACGCCATCTGCAGTATAGATGGATATGAT 369
 |||||

67 tHisArgLeuGlyCysProGlnAlaArgMetArgGluGluAsnMetGluA 84
 |||||

370 GCATAGGCTTGGAGAACCAAGCAAGGATGAGAGAGAGATATGGAAA 419
 |||||

84 rgileGlyGluGluValArgGlnLeuMetGluLysLeuArgGlyGln 100
 |||||

420 GGATTGGGAGGAGGTGAGACAGCTGATGGAAGCTGAGGGAAGCAG 469
 |||||

101 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspHi 117
 |||||

470 TTGAGTCATAGTCATGCGGGGAGTCAGCATCTGACCCCTCACCATGACCA 519
 |||||

117 sHisAspGluPheCysLeuMetPro 125
 |||||

520 TCATGATGAGTTTGCCTTATGCC 544

seq_name: gb_est1:AL526247

seq_documentation_block:

LOCUS AL526247 794 bp mRNA EST 13-FEB-2001
 DEFINITION LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC016Y001 5
 prime, mRNA sequence.

ACCESSION AL526247

VERSION AL526247.1 GI:12789740

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 794)

AUTHORS Li, W. B., Gruber, C., Jessee, J., and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

Location/Qualifiers

1..794

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="LTI_NFL003_NBC3"

/sex="male"

/tissue_type="neuroblastoma cells"

/lab_host="DH10B"

/note="organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
 was primed with a NotI-oligo(GT) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed
 by Life Technologies. Contact : Feng Liang Life
 Technologies, a division of Invitrogen 9800 Medical Center
 Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
 8371 Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com"

BASE COUNT 236 a 163 c 213 g 180 t

ORIGIN

alignment_scores:

Quality: 653.00 Length: 125

Ratio: 5.266 Gaps: 0

Percent Similarity: 99.200 Percent Identity: 98.400

alignment_block:

US-09-327-750D-32 x AL526247

Align seg 1/1 to: AL526247 from: 1 to: 794

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17

148 ATGGAGTCCAAAGAGAACGAGCAGTAAACAGTCTCAGCATGGAAATGC 197

17 aAsnGlnGluAsnGluGluLysGluGlnValAlaAsnLysGlyGluProL 34

198 CAACCAAGAAATGAAGAAGAGGACAGTTGCTAATAAAGGGAGCCCT 247

34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50

248 TGGCCCTCCCTTGGATGCTGGTGAATACCTGCTGCCTAGAGAAATCGT 297

51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTyrAspIleWe 67

298 AGCGCGTTCCCGTTCAGGAGCCCATCTCGAGTATAGATGGATATGAT 347

67 thisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 84

348 GCATAGGCTTGAGAACACACAGGAGGATGAGAGAGATATGAAA 397

84 rgIleGlyGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100

398 GGATTGGGAGAGGCTGAGACAGCTGATGGAAGAGCTGAGGAAAGCAG 447

101 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspH 117

448 TTGAGTCATAGTCTCGGGGAGTCTGAGCAGTACCCCTCACCATGACCA 497

117 shisAspGluPheCysLeuMetPro 125

498 TCATGATGAGTTTTCCTTATGCC 522

seq_name: gb_est1:AV702643

seq_documentation_block:

LOCUS AV702643 683 bp mRNA EST 08-OCT-2000

DEFINITION AV702643 ADB Homo sapiens cDNA clone ADBDBE11 5', mRNA sequence.

ACCESSION AV702643

VERSION AV702643.1 GI:10718973

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 683)

AUTHORS Peng, Y., Song, H., Huang, Q., Huang, C., Gu, Y., Yang, Y., Gao, G., Xiao,

H., Xu, X., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z.,

Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu,

G., Hu, R., Chen, J., Chen, Z., and Han, Z.

TITLE Homo sapiens cDNA ADB clones

JOURNAL Unpublished (2000)

COMMENT Contact: Zengqun Han

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919(ex.45)

Fax: 86-21-50801922

Email: hanzq@chgc.sh.cn

This clone is available at CHGC in Shanghai.

FEATURES

Location/Qualifiers

1..683

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="ADBDBE11"

/tissue_type="adrenal gland"

/dev_stage="Adult"

/lab_host="SOLR"

/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 208 a 128 c 184 g 161 t 2 others

ORIGIN

alignment_scores:

Quality: 650.00 Length: 125

Ratio: 5.242 Gaps: 0

Percent Similarity: 99.200 Percent Identity: 97.600

alignment_block:

US-09-327-750D-32 x AV702643

Align seg 1/1 to: AV702643 from: 1 to: 683

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17

60 ATGGAGTCCAAAGAGAACGAGCAGTAAACAGTCTCAGCATGGAAATGC 109

17 aAsnGlnGluAsnGluLysGluGlnValAlaAsnLysGlyGluProL 34

Email: cgaabs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10679 row: p column: 05
High quality sequence stop: 712.
High quality sequence stop: 712.

FEATURES
source
1. 734
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="479582"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site: 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.3 kb and
normalized to R05. This is a primary library enriched
for full-length clones and constructed using the
cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 187 a 174 c 214 g 159 t
ORIGIN

alignment_scores
Quality: 646.00 Length: 125
Ratio: 5.252 Gaps: 0
Percent Similarity: 98.400 Percent Identity: 97.600

alignment_block
US-09-327-750D-32 x BG715659
Align seg 1/1 to: BG715659 from: 1 to: 734

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||
206 ATGGAGTCCAAAGAGAAACGACGAGTAAACAGTCTCAGCATGGAATGCG 255
17 aasnGlnGluAsnGluLysGluLysValAlaAsnLysGlyGluProL 34
|||||
256 CACCAAGAAATGAGGAAAGGAGGAGTCTGCTAATAAGGGAGGCCCT 305
34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50
|||||
306 TGCCCTCCCTTTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 355
51 ArgArgPheProValArgGlnProLeuGlnTyrArgTyrPhePhe 67
|||||
356 AGCGGGTCCGGTTCGAGCAGCCCATCTGCGATATAGATGGATATGAT 405
67 tHisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 84
|||||
406 GCATAGGCTTGGAGAACCCACAGGAGGATGAGAGAGAGAAATGGA 455
84 rglGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100
|||||
456 GGATGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGGGAAGACG 505
101 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspHI 117
|||||
506 TTGAGTCATAGTCTCGGCGAGTACGAGTACGAGTACGAGTACGAGTAC 555
117 sHisAspGluPheCysLeuMetPro 125
|||||
556 TCATGATGAGTTTTCCTTATGCCC 580

seq_name: gb_est2:BG709325
seq_documentation_block: 828 bp mRNA EST 07-MAY-2001
LOCUS BG709325
DEFINITION 602673345F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:479582 5',
mRNA sequence.
ACCESSION BG709325
VERSION BG709325.1 GI:13987550
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 828)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10679 row: j column: 11
High quality sequence stop: 798.

FEATURES
source
1. 828
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="479582"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site: 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.3 kb and
normalized to R05. This is a primary library enriched
for full-length clones and constructed using the
cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 232 a 179 c 228 g 189 t
ORIGIN

alignment_scores
Quality: 637.00 Length: 126
Ratio: 5.137 Gaps: 1
Percent Similarity: 98.413 Percent Identity: 96.825

alignment_block
US-09-327-750D-32 x BG709325
Align seg 1/1 to: BG709325 from: 1 to: 828

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||
176 ATGGAGTCCAAAGAGAAACGACGAGTAAACAGTCTCAGCATGGAATGCG 225
17 aasnGlnGluAsnGluLysGluLysValAlaAsnLysGlyGluProL 34
|||||
226 CACCAAGAAATGAGGAAAGGAGGAGTCTGCTAATAAGGGAGGCCCT 275
34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50
|||||
276 TGCCCTCCCTTTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 325


```
7111 CTAGAGGGCCAGAAAAGAA.....AAGCAATACCACTAAAGTCTTT 7154
17 aasnGluAsnGluGluGluValAlaAsnLysGlyGluProL 34
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7155 AGAAGAAGAAAGTGTAGAGTCAGAGAAACTCTAGAAATGGGTTCTGT 7204
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
34 eualLeuProLeu.....AspAlaGlyGlu 42
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7205 TACTATCTGAGCTTTAGGAAAGAGACACAAGAACAGAGGATCAAGAA 7254
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
43 TyrCysValProArgGlyAsnArgArgPhe..... 53
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7255 TTAATGTCCTTAAGGTACACTAAGAGATTTTCATCTCTAGGAAGGA 7304
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
54 .....ProValArgGlnProIleLeuGln..... 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7305 AAGTCAAGAAGTAGTGAGGCTTCAAAAGAGGGGAACCTAGAAATCATGGA 7354
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
62 .....TyrArgTrpAspIleMetHisArgLeuGlyGluProGlnAlaArg 76
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7355 CAGCTTTTAAGAGGAGCCCAACCCACTGGGATTTCCAGGAGCTGAG 7404
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
76 ..... 76
7405 GACCAGATGCTTGAGAGACTGCTAGAGAAAGAGGATCAGAGCTTCCCAAG 7454
77 .....MetArgGluGluA 81
7455 GTCTCAGAGGAAGAGGACCAGGAGGCGATGTAGACCTCTGCAGAAAGAGA 7504
81 snMetGluArgIleGly.....GluGluValArgGlnLeuMetGluLys 95
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7505 ATCAGGAACCACTAGGCTATGAAGACGAGGAGGCGGCAGATACTTACAGAGA 7554
96 LeuArgGluLysGlnLeuSerHisSerLeuArgAla 107
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7555 CTATGAGAAAGAGAGTCCGAGGCTCTGAGGCT 7590
```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-306-691B-23

```
seq_documentation_block:
; Sequence 23, Application US/08306691B
; Patent No. 5734039
; GENERAL INFORMATION:
; APPLICANT: Calabretta, Bruno
; APPLICANT: Skorski, Tomasz
; TITLE OF INVENTION: ANTISENSE
; TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seidel, Gonda, Lavorgna & Monaco, P.C.
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/306,691B
; FILING DATE: September 15, 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A.
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-8
; TELECOMMUNICATION INFORMATION:
```

```
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; TELEX: No. 5734039e
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2301 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-306-691B-23

alignment_scores:
      Quality: 84.00      Length: 102
      Ratio: 1.355      Gaps: 4
      Percent similarity: 60.784      Percent identity: 32.353

alignment_block:
US-09-327-750D-32 x US-08-306-691B-23 ..
Align seg 1/1 to: US-08-306-691B-23 from: 1 to: 2301

      2 GluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAlaAs 18
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      77 GAGGAGCAGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 126
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      18 nGlnGluAsnGluGluLysGluValAlaAsnLysGlyGluProLeuA 35
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      127 GCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGCTG 176
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      35 lateuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArgArg 51
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      177 CAAGCCGAG.....CGGAGGAGG 195

      52 ArgPheProValArgGlnProIleLeu.....GlnTyrArgTTPAspI 66
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      196 CAGGACCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTCA 245
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      66 eMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMet 82
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      246 CCACCATCGAGGCGGTGAGCGCAAGATCCAGGTTCTGCAGCAGCAGGCA 295
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      83 GluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArgGluL 99
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      296 GATCATCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGA 345
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      99 ys 99
      : :
      346 AA 347
```

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-06251-78

```
seq_documentation_block:
; Sequence 78, Application PC/TUS9306251
; GENERAL INFORMATION:
; APPLICANT: Wickstrom, Eric and Rife, Jason P.
; TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
; TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: NY
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06251
```

```

; FILING DATE: 19930630
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Digilio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8586
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2301 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US93-06251-78

alignment_scores:
    Quality: 84.00      Length: 102
    Ratio: 1.355        Gaps: 4
    Percent Similarity: 60.784      Percent Identity: 32.353

alignment_block:
US-09-327-750D-32 x PCT-US93-06251-78
Align seg 1/1 to: PCT-US93-06251-78 from: 1 to: 2301

2 GluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAlaAs 18
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77 GAGGAGCAGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 126

18 nGlnGluAsnGluGluLysGluGlnValAlaAsnLysGlyGluProLeuA 35
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
127 GCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 176

35 laLeuProLeuAspLacGlyGluTyrCysValProArgGlyAsnArgArg 51
|||:|||||:|||||:|||||:|||||:|||||:|||||:
177 CAACGCCGAG.....CGGAGGAGG 195

52 ArgPheProValArgGlnProIleLeu.....GlnTyrArgTrpAspI 66
::: ||| |||:|||||:|||||:|||||:|||||:|||||:
196 CAGGAACCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 245

66 eMethHisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMet 82
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
246 CCACCATCGAGCGGTGAAGCGCAGATCCAGGTCTGCAGCAGCAGGCA 295

83 GluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArgGluL 99
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:
296 GATGATGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 345

99 ys 99
|||
346 AA 347

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-872-644-52
seq_documentation_block:
; Sequence 52, Application US/07872644
; Patent No. 5389527
; GENERAL INFORMATION:
; APPLICANT: Beavo, Joseph A.
; APPLICANT: Bentley, Kelley
; APPLICANT: Charbonneau, Harry
; APPLICANT: Sonnenburg, William K.
; TITLE OF INVENTION: DNA Encoding Mammalian
; TITLE OF INVENTION: Phosphodiesterases
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &

```

```

; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; STREET: Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/872,644
; FILING DATE: 19920420
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/688,356
; FILING DATE: 04-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5389527and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/30822
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2077 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: Single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1693
US-07-872-644-52

alignment_scores:
    Quality: 80.50      Length: 126
    Ratio: 1.118        Gaps: 6
    Percent Similarity: 57.143      Percent Identity: 24.603

alignment_block:
US-09-327-750D-32 x US-07-872-644-52
Align seg 1/1 to: US-07-872-644-52 from: 1 to: 2077

4 LysGluLysArgAlaValAsnSerLeuSerMetGluAsnAlaAsnGlnG 20
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
959 CAGAGGCGTTCGAGTTTGAATAGCATCAGCTCGTCAGATGCCAAGCATC 1008

20 u.....AsnGluGluLysGluGlnValAlaAsnLys. 30
|||||:|||||:|||||:|||||:|||||:|||||:
1009 AGGTGTCAAGACCTCTGGTTTCAGAGGAGTCCCGCATCAATCTG 1058

31 GlyGluProLeu.....AlaLeuProLeuAspAlaGlyGluTyrCy 44
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1059 TCATCTCCGTTGACTATAAGAGCTTTAAAGCTACTTTGGACGGAAGTGTG 1108

44 sval.....ProArgLysAsnArgArgPheP 54
|||||:|||||:|||||:|||||:|||||:|||||:
1109 CACATCAATCGGAGAGATGGAGCGCCAGAGTACCCTCAAGAGAGGAGGC 1158

54 roValArgGlnProIleLeuGlnTyrArgTrpAsp.....IleMethHis 68
|||:|||||:|||||:|||||:|||||:|||||:|||||:
1159 CAAGAAGGAAGCAGAGAGAAAGGCTCGCTGCGCCGACAGAGGAGCAGCAA 1208

69 ArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluArgI 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1209 AGGAATGGAAGCCAAAGCAGGCTGAAGAGGCGCATCTGGCAAGCT 1258

```

[illegible]

31 GlyGluProLeu.....AlaLeuProLeuAspAlaGlyGluTyrCy

```

101059 TCATCTCCGTTGACTATACAGAGCTTTAAAGCTACTTTGGACGGAAGTGGTG 1108
44 sval.....ProArgGlyAsnArgArgArgPhep 54
   :::          |          |          |          |          |
11111109 CACATCAATCGGAGAGATGGAGGGGCCAGGTACCCCAAAGAGAGAAGGC 1158
   |          |          |          |          |          |
54 roValArgGlnProIleLeuGlnTyArgTirPasp.....IleMetHis 68
   |          |          |          |          |          |
11111159 CAAGAAGAGACAGAGAAAGGCTGCCTGGCCGAGAGGAGCAGCAAA 1208
   |          |          |          |          |          |
69 ArgLeuGlyGluProGlnAlaArgMetArgGluGluGluAsnMetGluArgI 85
   |          |          |          |          |          |
11209 AGGAAATGGAAGCCAAAGCCAGGCTGAAGAAGGCGCATCTGGCAAAGCT 1258
   |          |          |          |          |          |
85 eGlyGluGluValArgGlnLeuMetGlnLysLeuArgGluLysGlnLeu 102
   |          |          |          |          |          |
11359 GAGA.....AAAGAGCGCTTGGAGAAACTTAAGATCAAGTCANTGG 1299
   |          |          |          |          |          |
102 erHisSerLeuArgAlaValSerThr 110
   |          |          |          |          |          |
11300 AACACGGGCAACAAAAGTGCACAACC 1325
   |          |          |          |          |          |

name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:us-08-297-510-52

documentation_block:
sequence 52, Application US/08297510
tent No. 5602019
GENERAL INFORMATION:
APPLICANT: Beavo, Joseph A.
APPLICANT: Bentley, Kelley
APPLICANT: Charbonneau, Harry
APPLICANT: Sonnenberg, William K.
TITLE OF INVENTION: DNA Encoding Mammalian
Phosphodiesterases
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESS: Marshall, O'Toole, Gerstein, Murray &
ADDRESS: Bicknell
STREET: Two First National Plaza, 20 South Clark
STREET: Street
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/297,510
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/688,356
FILING DATE: 04-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: No. 5602019and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/30822
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 346-5750

```

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/688,356
FILING DATE: 04-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: NO. 5602019and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/30822
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 346-5750

NAME: NO. 3602019and, Gilda E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/30822.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 346-5750

TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 346-5750

```
TELEFAX: (312) 984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 2077 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 2..1693
US-08-297-510-52

alignment_scores:
Quality: 80.50 Length: 126
Ratio: 1.118 Gaps: 6
Percent Similarity: 57.143 Percent Identity: 24.603

alignment_block:
US-09-327-750D-32 x US-08-297-510-52

Align seg 1/1 to: US-08-297-510-52 from: 1 to: 2077

4 LysGluLysArgAlaValAsnSerLeuSerMetGluAsnAlaAsnGlnG1 20
|||||
959 CAGAGCGCTTCGAGTTGTAATAGCATCAGCTCGTCAGATGCCAAGCGATC 1008
20 u.....AsnGluGluLysGluGlnValAlaAsnLys. 30
1009 AGGTGTCAGACCTCTGTTTCAGAGGGAAGTCCCGCATCAACAATTCG 1058
31 GlyGluProLeu.....AlaLeuProLeuAspAlaGlyGluTyrCy 44
|||||
1059 TCATCTCCGTTGACTATAAGAGCTTTAAAGCTACTTGGACGGAAGTGGTG 1108
44 sVal.....ProArgGlyAsnArgArgPheP 54
|||||
1109 CACATCAATCGGAGATGGAGGCCAAGGTACCCCAAGAGGAGGAAGGC 1158
54 roValArgGlnProIleLeuGlnTyrArgTrpAsp.....IleMethis 68
|||||
1159 CAAGAAGCAAGCAGAGAAAGGCTCGCTGGCGCAGAGGAGCAGCAAA 1208
69 ArgLeuGlyGluProGlnAlaArgMetArgGluAsnMetGluArgil 85
|||||
1209 AGGAATGGAAGCCAAAGCCAGCTGAAGAAGCGCATCTGGCAAGCT 1258
85 eGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGlnLeus 102
|||||
1259 GAGA.....AAAGACGCTCTGGAGAACTAAGAATCAAGTCAATGG 1299
102 erHisSerLeuArgAlaValSerThr 110
|||||
1300 AACACGGGCAACAAAGTGACAAACC 1325

seq_name: /cgn2_5/ptodata/2/ina/5A_COMB.seq:US-08-479-532-52

seq_documentation_block:
Sequence 52, Application US/08479532
Patent No. 5776752
GENERAL INFORMATION:
APPLICANT: Beavo, Joseph A.
APPLICANT: Bentley, Kelley
APPLICANT: Charbonneau, Harry
APPLICANT: Sonnenburg, William K.
TITLE OF INVENTION: DNA Encoding Mammalian
TITLE OF INVENTION: Phosphodiesterases
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSEE: Bicknell
```

1209 AGGAAATGGAAGCCAAAGCCAGGCTGAAGAAGGCGCATCTGGCAAAGCT 1258
85 eGlyGluValArgGlnLeuMetGluLysLeuArgGluLysGlnLeuS 102
1259 GAGA.....AAAGACGCTGGGAAATTAAGATCAAGTCAATGG 1299
102 erHisSerLeuArgAlaValSerThr 110
1300 AACACGGGCAACAAAGTGACAACC 1325
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-455-526-52

seq_documentation_block:
; Sequence 52, Application US/08455526
; Patent No. 5789553
; GENERAL INFORMATION:
; APPLICANT: Beavo, Joseph A.
; APPLICANT: Bentley, Kelley
; APPLICANT: Charbonneau, Harry
; APPLICANT: Sonnenburg, William K.
; TITLE OF INVENTION: DNA Encoding Mammalian
; TITLE OF INVENTION: Phosphodiesterases
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/455,526
; FILING DATE: 31-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/297,494
; FILING DATE: 29-AUG-1994
; APPLICATION NUMBER: US 07/688,356
; FILING DATE: 04-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5789553and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/30822
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2077 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1693
US-08-455-526-52

alignment_scores:
Quality: 80.50 Length: 126
Ratio: 1.118 Gaps: 6
Percent Similarity: 57.143 Percent Identity: 24.603
alignment_block:

US-09-327-750D-32 x US-08-455-526-52
Align seg 1/1 to: US-08-455-526-52 from: 1 to: 2077
4 LysGluLysArgAlaValAsnSerLeuSerMetGluAsnAlaAsnGlnCl 20
959 CAGAGGCGTTCGAGTTTGAATAGCATCAGCTCGTCAGATGCCAAGCATC 1008
20 u.....AsnGluLysGluGlnValAlaAsnLys. 30
1009 AGGTGTCAAGACCTCTGTTTCAGAGGAAGTCCCGCATCAACAATTCGTG 1058
31 GlyGluProLeu.....AlaLeuProLeuAspAlaGlyGluTyCy 44
1059 TCATCTCCGTTGACTATAAGAGCTTTAAAGCTACTTTGGAGCGGAAGTGGTG 1108
44 sVal.....ProArgGlyAsnArgArgPheP 54
1109 CACATCAATCGGGAGAGATGGAGGCCAAGGTACCCAAAGAGAGAGAGGC 1158
54 roValArgGlnProIleLeuGlnTyArgTirPasp.....IleMetHis 68
1159 CAAGAAGGAGCAGAGGAAAGGCTCGCTGCGCCGAGAGAGAGAGCAGCAA 1208
69 ArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluArgI 85
1209 AGGAAATGGAAGCCAAAGCCAGGCTGAAGAAGGCGCATCTGGCAAAGCT 1258
85 eGlyGluValArgGlnLeuMetGluLysLeuArgGluLysGlnLeuS 102
1259 GAGA.....AAAGACGCTGGGAAATTAAGATCAAGTCAATGG 1299
102 erHisSerLeuArgAlaValSerThr 110
1300 AACACGGGCAACAAAGTGACAACC 1325
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-455-526-52

seq_documentation_block:
; Sequence 52, Application US/08455525
; Patent No. 5800987
; GENERAL INFORMATION:
; APPLICANT: Beavo, Joseph A.
; APPLICANT: Bentley, Kelley
; APPLICANT: Charbonneau, Harry
; APPLICANT: Sonnenburg, William K.
; TITLE OF INVENTION: DNA Encoding Mammalian
; TITLE OF INVENTION: Phosphodiesterases
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/455,525
; FILING DATE: 31-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/297,494
; FILING DATE:
; APPLICATION NUMBER: US 07/688,356
; FILING DATE: 04-APR-1991
; ATTORNEY/AGENT INFORMATION:

us-09-327-750d-32.rni

Tue Mar 12 09:01:37 2002

```
;
; NAME: No. 5800987and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/30822
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2077 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1693
; US-08-455-525-52

alignment_scores:
    Quality: 80.50      Length: 126
    Ratio: 1.118      Gaps: 6
    Percent Similarity: 57.143      Percent Identity: 24.603

alignment_block:
US-09-327-750D-32 x US-08-455-525-52
..
Align seg 1/1 to: US-08-455-525-52 from: 1 to: 2077

4 LysGluLysArgAlaValAsnSerLeuSerMetGluAsnAlaAsnGlnG1 20
959 CAGAGCGGTTTCGAGTTGAATAGCATCAGCTCGTCAGATGCCAAGCGATC 1008
20 u.....AsnGluLysGluGlnValAlaAsnLys. 30
1009 AGGTGTCAGACCTCTGGTTTCAGAGGGAAGTCCCGCATCAACAATTCG 1058
31 GlyGluProLeu.....AlaLeuProLeuAspAlaGlyGluTyrCy 44
1059 TCATCTCCGTTGACTATAAGAGCTTTAAAGCTACTTGGACGGAAGTGGTG 1108
44 sVal.....ProArgGlyAsnArgArgArgPhep 54
1109 CACATCAATCGGAGAGATGGAGGCCCAAGGTACCCAAAGAGGAGAGGC 1158
54 roValArgGlnProLeuGlnTyrArgTrpasp.....IleMetHis 68
1159 CAGAAGGAGGAGAGGAGGAAAGGCTCGCTGGCCGCGCAGAGGAGCAGCAA 1208
69 ArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluArgil 85
1209 AGGAATGGAAGCCAAAGCCAGGCTGAAGAAGCGCATCTGCGCAAGCT 1258
85 eglyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGlnLeus 102
1259 GAGA.....AAAGACCTCTGGAAACTAAGAATCAAGTCAATGG 1299
102 exHisSerLeuArgAlaValSerThr 110
1300 AACACGGGCAACAAACAAAGTGACAAC 1325

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-139-491-52

seq_documentation_block:
; Sequence 52, Application US/09139491
; Patent No. 6015677
; GENERAL INFORMATION:
; APPLICANT: Beavo, Joseph A.
; APPLICANT: Bentley, Kelley
; APPLICANT: Charbonneau, Harry
; APPLICANT: Sonnenburg, William K.
; TITLE OF INVENTION: DNA Encoding Mammalian
```



```

; FILING DATE: 04-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: NO. 5389527and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/30822
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2693 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 176..2077
; US-07-872-644-50

alignment_scores:
  Quality: 80.50 Length: 126
  Ratio: 1.118 Gaps: 6
  Percent Similarity: 57.143 Percent Identity: 24.603

alignment_block:
US-09-327-750D-32 x US-07-872-644-50 ..
Align seg 1/1 to: US-07-872-644-50 from: 1 to: 2693

4 LysGluLysArgAlaValAsnSerLeuSerMetGluAsnAlaAsnGlnG1 20
1568 CAGAGCGCTTCGAGTTGAATAGCATCAGCTCGTCAGATGCCAAGCGATC 1617
20 u.....AsnGluGluLysGluGlnValAlaAsnLys. 30
1618 AGGTCTCAAGACCTCTGGTTCAGAGGGAAGTCCCGCATCAACATCTG 1667
31 GlyGluProLeu.....AlaLeuProLeuAspAlaGlyGluTyrCy 44
1668 TCATCTCCGTTGACTATAAGAGCTTTAAAGCTACTTGGACGGAAGTGGTG 1717
44 sval.....ProArgGlyAsnArgArgArgPhep 54
1718 CACATCAATCGGAGAGATGGAGGCCCAAGGTACCCAAAGAGGAGGAGGC 1767
54 roValArgGlnProLeuGlnTyrArgTrpAsp.....IleMetHis 68
1768 CAAGAAGGAAGCAGAGGAAAGGCTCGCTGGCGCAGAGGAGCAGCAAA 1817
69 ArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluArg1 85
1818 AGGAATGGAAGCCAAAGCAGGCTGAAGAGCGGCATCTGCCAAGCT 1867
85 eGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGlnLeu 102
1868 GAGA.....AAAGACGCTGGAGAAACTAAGAATCAAGTCAATGG 1908
102 eHisSerLeuArgAlaValSerThr 110
1909 AACACGGGCAACAAAGTGACACC 1934

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-297-494-50
seq_documentation_block:
; Sequence 50: Application US/08297494
; Patent No. 5580771
; GENERAL INFORMATION:
; APPLICANT: Beavo, Joseph A.
; APPLICANT: Bentley, Kelley
; APPLICANT: Charbonneau, Harry

```

```

; APPLICANT: Sonnenburg, William K.
; TITLE OF INVENTION: DNA Encoding Mammalian
; TITLE OF INVENTION: Phosphodiesterases
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; STREET: Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/297,494
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/688,356
; FILING DATE: 04-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 558077land, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/30822
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2693 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 176..2077
; US-08-297-494-50

alignment_scores:
  Quality: 80.50 Length: 126
  Ratio: 1.118 Gaps: 6
  Percent Similarity: 57.143 Percent Identity: 24.603

alignment_block:
US-09-327-750D-32 x US-08-297-494-50 ..
Align seg 1/1 to: US-08-297-494-50 from: 1 to: 2693

4 LysGluLysArgAlaValAsnSerLeuSerMetGluAsnAlaAsnGlnG1 20
1568 CAGAGCGCTTCGAGTTGAATAGCATCAGCTCGTCAGATGCCAAGCGATC 1617
20 u.....AsnGluGluLysGluGlnValAlaAsnLys. 30
1618 AGGTCTCAAGACCTCTGGTTCAGAGGGAAGTCCCGCATCAACATCTG 1667
31 GlyGluProLeu.....AlaLeuProLeuAspAlaGlyGluTyrCy 44
1668 TCATCTCCGTTGACTATAAGAGCTTTAAAGCTACTTGGACGGAAGTGGTG 1717
44 sval.....ProArgGlyAsnArgArgArgPhep 54
1718 CACATCAATCGGAGAGATGGAGGCCCAAGGTACCCAAAGAGGAGGAGGC 1767
54 roValArgGlnProLeuGlnTyrArgTrpAsp.....IleMetHis 68

```

```

1768 CAAGAAGGAAGCAGGAGAAAGCTGCCTGCCCGCAGAGGACGACCAA 1817
69 ArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluArgI1 85
1818 AGGAATGAAGCCAAAGCCAGGCTCAAGAGCGCCATCTGGCAAGCT 1867
85 eGlyGluValArgGlnLeuMetGluLysLeuArgGluLysGlnLeu 102
1868 GAGA.....AAAAGCTCTGGAGAACTAAGATCAAGTCAATGG 1908
102 erHisSerLeuArgAlaValSerThr 110
1909 AACAGGGCAACAAAGTGACAACC 1934

```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-297-510-50

```

seq_documentation_block:
; Sequence 50, Application US/08297510
; Patent No. 5602019
; GENERAL INFORMATION:
; APPLICANT: Beavo, Joseph A.
; APPLICANT: Bentley, Kelley
; APPLICANT: Charbonneau, Harry
; APPLICANT: Sonnenburg, William K.
; TITLE OF INVENTION: DNA Encoding Mammalian
; TITLE OF INVENTION: Phosphodiesterases
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; STREET: Two First National Plaza, 20 South Clark
; CITY: Street
; STATE: Chicago
; CITY: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/297,510
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/688,356
; FILING DATE: 04-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5602019and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/30822
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2693 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 176..2077
US-08-297-510-50

```

alignment_scores: Quality: 80.50 Length: 126
Ratio: 1.118 Gaps: 6

```

Percent Similarity: 57.143 Percent Identity: 24.603
alignment_block:
US-09-327-750D-32 x US-08-297-510-50
Align seg 1/1 to: US-08-297-510-50 from: 1 to: 2693
4 LysGluLysArgAlaValAsnSerLeuSerMetGluAsnAlaAsnGlnG1 20
1568 CAGAGCGCTTCGAGTTTGAATAGCATCAGCTCGTCAGATGCCAAGGATC 1617
20 u.....AsnGluGluLysGluGlnValAlaAsnLys. 30
1618 AGGTGTCAGACCTCTGCTTCAGAGGAAGTCCCGCATCAACAATTCG 1667
31 GlyGluProLeu.....AlaLeuProLeuAspAlaGlyGluTyrCy 44
1668 TCATCTCGTTGACTATAAGAGCTTTAAAGCTACTTGGACGGAAGTGGTG 1717
44 sVal.....ProArgGlyAsnArgArgArgPhep 54
1718 CACATCAATCGGAGAGATGGAGGGCCAGGTACCCAAAGAGGAGAAGGC 1767
54 roValArgGlnProIleLeuGlnTyrArgTirPasp.....lleMetHis 68
1768 CAAGAAGGAAGCAGAGGAAAGGCTCGCTGCCCGCAGAGGACGACCAA 1817
69 ArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluArgI1 85
1818 AGGAATGAAGCCAAAGCCAGGCTCAAGAGCGCCATCTGGCAAGCT 1867
85 eGlyGluValArgGlnLeuMetGluLysLeuArgGluLysGlnLeu 102
1868 GAGA.....AAAAGCTCTGGAGAACTAAGATCAAGTCAATGG 1908
102 erHisSerLeuArgAlaValSerThr 110
1909 AACAGGGCAACAAAGTGACAACC 1934
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-479-532-50

```

```

seq_documentation_block:
; Sequence 50, Application US/08479532
; Patent No. 5776752
; GENERAL INFORMATION:
; APPLICANT: Beavo, Joseph A.
; APPLICANT: Bentley, Kelley
; APPLICANT: Charbonneau, Harry
; APPLICANT: Sonnenburg, William K.
; TITLE OF INVENTION: DNA Encoding Mammalian
; TITLE OF INVENTION: Phosphodiesterases
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; STREET: Two First National Plaza, 20 South Clark
; CITY: Street
; STATE: Chicago
; CITY: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,532
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/297,494
; FILING DATE:

```

Tue Mar 12 09:01:37 2002

```

APPLICATION NUMBER: US 07/688,356
FILING DATE: 04-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: No. 5776752and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/30822
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 346-5750
TELEFAX: (312) 984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 2693 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 176..2077
US-08-479-532-50

alignment_scores:
Quality: 80.50 Length: 126
Ratio: 1.118 Gaps: 6
Percent Similarity: 57.143 Percent Identity: 24.603

alignment_block:
US-09-327-750D-32 x US-08-479-532-50

Align seg 1/1 to: US-08-479-532-50 from: 1 to: 2693

4 LysGluLysArgAlaValAsnSerLeuSerMetGluAsnAlaAsnGlnG1 20
1568 CAGAGCGTTCGAGTTTGAATAGCATCAGCTGTCAGATGCCAAGCGATC 1617
20 u.....AsnGluGluLysGluGlnValAlaAsnLys. 30
1618 AGGTGTCAAGACCTCTGGTTCAGAGGGAAGTGGCCCGCATCAACAATCTG 1667
31 GlyGluProLeu.....AlaLeuProLeuAspAlaGlyGluTyrCy 44
1668 TCATCTCCGTTGACTATAGAGCTTTAAAGCTACTTGGACGGAAGTGGTG 1717
44 sVal.....ProArgGlyAsnArgArgPheP 54
1718 CACATCAATCGGAGAGATGGAGGCCCAAGGTACCCAAAGAGGAGAGGC 1767
54 roValArgGlnProIleLeuGlnTyrArgTyrPasp.....IleMethHis 68
1768 CAAGAAGGAAGCAGAGGAAAGGCTCGCTGGCCGACAGAGGAGCAGCAA 1817
69 ArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluArgI1 85
1818 AGGAATGGAGCAACCAAGCCAGGCTGAAGAGCGGCATCTGGCAAGCT 1867
85 eGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGlnLeus 102
1868 GACA.....AAAAGACGCTCTGGAGAACTAAGAATCAAGTCAATGG 1908
102 erHisSerLeuArgAlaValSerThr 110
1909 AACACGGGCAACAAAGTGCACACC 1934
```

2683	/SIDS2/gcgdata/geneseq/geneseq/NAL198	DAT:AAV54765	+	80.50	136.79	15.27
2684	/SIDS2/gcgdata/geneseq/geneseq/NAL198	DAT:AAV48208	+	80.50	136.79	15.27
2685	/SIDS2/gcgdata/geneseq/geneseq/NAL198	DAT:AAV36160	+	80.50	136.79	15.27
2686	/SIDS2/gcgdata/geneseq/geneseq/NAL2000	DAT:AAZ30384	+	80.50	136.79	15.27
2687	/SIDS2/gcgdata/geneseq/geneseq/NAL1999	DAT:AAZ20086	+	80.50	124.99	10.07
2688	/SIDS2/gcgdata/geneseq/geneseq/NAL1999	DAT:AAZ20086	+	80.50	124.99	10.07

seq_name: /SIPDS2/acqdata/qeneseq/qeneseqn/NA2001.DAT:AAH45143

2/app_query.fasta_1.1472

Sequence	Strd Orig	zScore	EScore Len	Documenta-
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT.AAH45143	+	639.00	1290.8	639.00
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT.AAH75810	+	602.50	1216.5	602.50
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT.AAH58591	+	602.50	1215.3	602.50
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT.AAF59611	+	602.50	1214.9	602.50
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT.AAF160367	+	589.50	1188.8	589.50
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT.AAC03880	+	585.50	1183.2	585.50
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT.AAC03879	+	585.50	1182.7	585.50
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT.AAC01889	+	562.50	1135.7	562.50
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT.AAC06183	+	207.50	414.9	207.50
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT.AAC06180	+	207.50	414.1	207.50
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT.AAH03517	+	202.50	397.5	202.50
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT.AAH13750	+	202.50	394.0	202.50
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT.AAC85548	+	202.50	393.0	202.50
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT.AAF23529	+	201.00	394.1	201.00
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT.AAF21748	+	185.50	362.2	185.50
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT.AAF32528	+	185.00	367.2	185.00
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT.AAC06184	+	184.50	367.2	184.50
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT.AAC01005	+	151.50	298.1	151.50
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT.AAC01004	+	136.50	268.3	136.50
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT.AAC06181	+	117.00	234.0	117.00
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT.AAH87107	+	106.00	207.9	106.00
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT.AAH87143	+	100.50	204.3	100.50
/SID22/gcgdata/geneseq/geneseq/NA198.DAT.AAH14053	+	89.00	160.2	89.00
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT.AAH82586	+	88.50	171.5	88.50
/SID22/gcgdata/geneseq/geneseq/NA192.DAT.AAC028398	+	86.50	141.7	86.50
/SID22/gcgdata/geneseq/geneseq/NA194.DAT.AA070447	+	86.50	134.1	86.50
/SID22/gcgdata/geneseq/geneseq/NA198.DAT.AAV20943	+	84.00	145.5	84.00
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT.AAH509303	+	84.00	145.5	84.00
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT.AAH15730	+	81.00	135.1	81.00
/SID22/gcgdata/geneseq/geneseq/NA193.DAT.AAQ30188	+	80.50	139.3	80.50
/SID22/gcgdata/geneseq/geneseq/NA195.DAT.AAQ3982	+	80.50	139.3	80.50
/SID22/gcgdata/geneseq/geneseq/NA197.DAT.AAT67203	+	80.50	139.3	80.50
/SID22/gcgdata/geneseq/geneseq/NA197.DAT.AAT51117	+	80.50	139.3	80.50
/SID22/gcgdata/geneseq/geneseq/NA198.DAT.AAV54766	+	80.50	139.3	80.50
/SID22/gcgdata/geneseq/geneseq/NA198.DAT.AAV484209	+	80.50	139.3	80.50
/SID22/gcgdata/geneseq/geneseq/NA198.DAT.AAV36161	+	80.50	139.3	80.50
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT.AAQ30385	+	80.50	139.3	80.50
/SID22/gcgdata/geneseq/geneseq/NA195.DAT.AAQ3981	+	80.50	136.7	80.50
/SID22/gcgdata/geneseq/geneseq/NA197.DAT.AAT2202	+	80.50	136.7	80.50
/SID22/gcgdata/geneseq/geneseq/NA197.DAT.AAT51116	+	80.50	136.7	80.50

alignment_scores:			
Quality:	639.00	Length:	125
Ratio:	5.238	Gaps:	0
Percent Similarity:	97.500	Percent Identity:	96.800

```
alignment_block:
US-09-327-750D-32 x AAH45143 ..
```

MetClausenCliffordArchaivaldenSerMetClifford 17

173 ATGGAGTCCAAAGAGAGAACTAGCAGTAAACAGTCTCAGCATGGAAAATGC 222

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

```

34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50
|||||
273 TGGCCCTCCCTTTGGATGCTGGTGAATACGTGTGCTAGAGGAATCGT 322
|||||
51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTyrAspIleMe 67
|||||
323 AGGCGGTTCCTCCGTTAGCGACCCCATCTGCAGTATAGATGGATGAT 372
|||||
67 HisArgLeuGlyGluProGlnIleArgMetArgGluAsnMetGluA 84
|||||
373 GCATAGGCTTGAGAACACACAGCAGGATGAGAGAGATATATGGAAA 422
|||||
84 TgIleGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100
|||||
423 GCATTTGGGGAGGGGTGAGACAGCTGATGAAAAGCTGAGGAAAAGCAG 472
|||||
101 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspHi 117
|||||
473 TTGAGTCATAGCTGCGGGCAGTCAGCACTGACCCCTCACCATGACCA 522
|||||
117 shiAspGluPheCysLeuMetPro 125
|||||
523 TCATGATGAGTTTTCCTTATGCC 547
|||||

```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH75810

seq_documentation_block:
ID AAH75810 standard; cdna; 767 BP.

AC AAH75810;

DT 17-OCT-2001 (first entry)

DE Human X chromosome linked gene expression protein 14 coding sequence.

XX Human; X chromosome linked gene expression protein 14; cancer;

KW HIV infection; cytostatic; anti-HIV; chromosome X; ss.

XX Homo sapiens.

OS CN1296969-A.

XX 30-MAY-2001.

XX 23-NOV-1999; 99CN-0124078.

XX 23-NOV-1999; 99CN-0124078.

XX (SHAN-) SHANGHAI BORONG GENE DEV CO LTD.

PI Mao Y, Xie Y;

XX WPI: 2001-483897/53.

DR P-PSDB; AAG66407.

XX Polypeptide-human X chromosome linked gene expression protein 14 and polynucleotide for coding said polypeptide.

XX Claim 6; Page 24 (Disclosure); 31pp; Chinese.

CC The present sequence is the coding sequence for human X chromosome linked gene expression protein 14. The protein and coding sequence are useful for treating diseases e.g. cancer and HIV infection.

XX Sequence 767 BP; 224 A; 143 C; 215 G; 185 T; 0 other;

alignment_scores:

Quality: 602.50

Length: 128

Ratio: 4.898

Gaps: 1

Percent Similarity: 96.094

Percent Identity: 89.062

alignment_block:

US-09-327-750D-32 x AAH75810

Align seg 1/1 to: AAH75810 from: 1 to: 767

```

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||
101 ATGGAGTCCAAAGAGGAACGAGCGTTAAACAATCTCATCGTGGAAAATG 150
|||||
17 aasnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31
|||||
151 CAACCCAGGAAATGATGAAAAGATGAAAAGGAGCAAGTTGCTAATAAAG 200
|||||
31 lyGluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47
|||||
201 GGGAGCCCTTGGCCCTACCTTTGAATGTTAGTGAATACTGTGTGCCTAGA 250
|||||
48 GlyAsnArgArgPheProValArgGlnProIleLeuGlnTyrArgTr 64
|||||
251 GGAACCCGTAGCGGTTCCGCTTAGCGCCCATCTGCAGTATAGATG 300
|||||
64 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluAla 81
|||||
301 GGACATAATGTCATAGCTTGGAGAGCCACAGCAAGGATGAGAGAGGAGA 350
|||||
81 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 97
|||||
351 ATATGGAAGGATTTGGGGAGGAGGTGAGACAGCTGATGGAAAAGCTGAGG 400
|||||
98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 114
|||||
401 GAAAAGCAGATTGAGTCATAGTCTCGGGCAGTCAGCACACTGATCCCCCTCA 450
|||||
114 shiAspHisHisAspGluPheCysLeuMetPro 125
|||||
451 CCATGACCATCATCAGTATGAGTTTTCCTTATGCC 484
|||||

```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI58581

seq_documentation_block:

ID AAI58581 standard; cdna; 862 BP.

XX AAI58581;

XX 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 784.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia; ss.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX WPI: 2001-442253/47.
 DR P-PSDB; AAM39425.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX Claim 1; SEQ ID NO 784; 10078pp; English.
 PS
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic, and
 CC immunosuppressant and cyostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 862 BP; 235 A; 183 C; 238 G; 206 T; 0 other;

alignment_scores:
 Quality: 602.50 Length: 128
 Ratio: 4.898 Gaps: 1
 Percent Similarity: 96.094 Percent Identity: 89.062
 alignment_block:
 US-09-327-750D-32 x AAI58581
 Align seg 1/1 to: AAI58581 from: 1 to: 862
 1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
 226 ATGGAGTCCAAAGAGGACGAGCGTTAAACAATCTCATCTCGTGAATGCT 275
 17 abasnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31
 276 CAACAGGAAATGATGAAAAAGATGAAAGAGCAAGTTGCTTAATAAG 325
 31 lYGluproLeuAlaLeuProLeuAspAlaGlyGluTyrcysValProArg 47
 326 GGGACCCCTGGCCCTACCTTTGATGTTAGTACTGATGATGCTGCTAGA 375
 48 GlyAsnArgArgPheProValArgGlnProIleLeuGlnTyrrargtr 64
 376 GGAACCCCTAGCGGTTCGCGTTAGGACGCCATCCCTGCGATATAGATG 425
 64 pAspIleMetHisArgLeuGluGluProGlnAlaArgMetArgGluGlu 81
 426 GCACATATGATAGCTTGGAGAGCCACAGGCAAGGATGAGAGAGAGA 475
 81 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 97
 476 ATATGGAAGAGGATGGGAGGAGGTGAGACAGCTGATGGAAAGCTGAGG 525
 98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHI 114
 526 GAAGAGGAGTGTGATCATAGTTTCGGGCGAGTCACTGATCCCTCTCA 575
 114 shIsAspHisAspGluPheCysLeuMetPro 125
 ||||||||||||||||||||||||||||||||||||||||||||

576 CCATGACCATCAGATGAGTTTTCGCTTATGCC 609

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAF59611

seq_documentation_block:
 ID_AAF59611 standard; cDNA; 898 BP.
 AC AAF59611;
 XX 24-APR-2001 (first entry)
 DT XX
 DE Human cell cycle and proliferation protein CCYPR-22 cDNA, SEQ ID NO:76.
 XX Cell cycle and proliferation protein; CCYPR; human; agonist;
 KW antagonist; gene therapy; detection; gene therapy;
 KW transgenic animal disease model; immune disorder;
 KW developmental disorder; cell signalling disorder;
 KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;
 KW arteriosclerosis; asthma; allergy; diabetes mellitus;
 KW menstrual cycle disorder; bacterial infection; ss.
 XX Homo sapiens.
 OS
 XX WO200107471-A2.
 PN
 XX 01-FEB-2001.
 PD
 XX 21-JUL-2000; 2000WO-US19948.
 PF
 XX 21-JUL-1999; 99US-0145075.
 PR
 XX 08-SEP-1999; 99US-0153129.
 PR
 XX 10-NOV-1999; 99US-016447.
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 XX Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;
 PI Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;
 XX WPI: 2001-112727/12.
 DR P-PSDB; AAB60474.
 XX
 PT Human cell cycle and proliferation proteins and polynucleotides are
 PT used to treat, diagnose and prevent immune, developmental and cell
 PT signaling disorders and cell proliferative disorders including cancer -
 PS Claim 5; Page 181-182; 205pp; English.
 XX
 CC Sequences AAF59590-AAF59643 represent cDNAs encoding 54 human
 CC cell cycle and proliferation proteins (CCYPR), AAB60453-AAB60506.
 CC CCYPR and agonists of CCYPR are used to treat diseases or conditions
 CC associated with decreased expression of functional CCYPR, while CCYPR
 CC antagonists are used to treat diseases or conditions associated with
 CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies
 CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or
 CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect
 CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)
 CC that specifically bind to CCYPR, and in drug screening methods to
 CC identify compounds that modulate the activity of CCYPR. CCYPR
 CC nucleotides can be used to generate transgenic animal models of human
 CC disease, and can be used in gene therapy in target cells with genetic
 CC abnormalities with respect to the expression of CCYPR for the
 CC treatment or prevention of a disorder associated with CCYPR.
 CC Diseases which can be diagnosed, treated and prevented using CCYPR
 CC proteins, nucleic acids, agonists or antagonists include immune,
 CC developmental and cell signalling disorders, and cell proliferative
 CC disorders including cancer. Specific examples of these disorders
 CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
 CC diabetes mellitus, disorders of the menstrual cycle and infections
 CC caused by bacteria.
 XX
 SQ Sequence 898 BP; 250 A; 186 C; 251 G; 211 T; 0 other;

alignment_scores:

Quality: 602.50 Length: 128
 Ratio: 4.898 Gaps: 1
 Percent Similarity: 96.094 Percent Identity: 89.062

alignment_block:

US-09-327-750D-32 x AAF59611

Align seg 1/1 to: AAF59611 from: 1 to: 898

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
 |||||
 254 ATGGAGTCCAAAGAGGAACGAGCGTTAAACAATCTCATCGTGGAATGT 303
 |||||
 17 aAsnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31
 |||||
 304 CAACCGAGAAATGATGAAGAAAGATGAAAGGAGCAAGTTGCTAATAAG 353
 |||||
 31 lyGluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47
 |||||
 354 GGGAGCCCTTGGCCCTACCTTGAATGTTAGTACTGTGTGCTAGA 403
 |||||
 48 GlyAsnArgArgPheProValArgGlnProIleLeuGlnTyrArgTr 64
 |||||
 404 GGAACCGTAGGGGTTCCGCGTTAGGCAGCCCATCTCGAGTATAGATG 453
 |||||
 64 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGlu 81
 |||||
 454 GGACATATCATAGCTTGGAGAGCCACAGCAGGATGAGAGGAGA 503
 |||||
 81 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 97
 |||||
 504 ATATGGAAGGATTGGGGAGGAGGTGAGACAGCTGATGGAAGCTGAGG 553
 |||||
 98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 114
 |||||
 554 GAAAGCAGTTGAGTCATAGTCTGCGGGCAGTCAGCAGTATCCCTCA 603
 |||||
 114 sHisAspHisHisAspGluPheCysLeuMetPro 125
 |||||
 604 CCATGACCATCAGCATGAGTTTGCCTTATGCCC 637

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAI60367

seq_documentation_block:

ID: AAI60367 standard; cDNA; 858 BP.

XX AAI60367;

XX 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 4356.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.

XX Homo sapiens.

XX W0200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

PR

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

DR P-PSDB; AAM41211.

XX Novel nucleic acids and polypeptides, useful for treating disorders

PT such as central nervous system injuries -

XX Claim 1: SEQ ID NO 4356; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and

CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,

CC immunosuppressant and cytostatic activity. The polynucleotides are useful

CC in gene therapy. A composition containing a polypeptide or polynucleotide

CC of the invention may be used to treat diseases of the peripheral nervous

CC system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

CC utilisation of the activities such as: Immune system suppression,

CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

CC assays for receptor activity, arthritis and inflammation, leukaemias and

CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed

CC specification.

XX Sequence 858 BP; 236 A; 175 C; 241 G; 206 T; 0 other;

alignment_scores:

Quality: 589.50 Length: 129

Ratio: 4.793 Gaps: 2

Percent Similarity: 95.349 Percent Identity: 88.372

alignment_block:

US-09-327-750D-32 x AAI60367

Align seg 1/1 to: AAI60367 from: 1 to: 858

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17

|||||

215 ATGGAGTCCAAAGAGGAACGAGCGTTAAACAATCTCATCGTGGAATGT 264

|||||

17 aAsnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31

|||||

265 CAACCGAGAAATGATGAAGAAAGATGAAAGGAGCAAGTTGCTAATAAG 314

|||||

31 lyGluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47

|||||

315 GGGAGCCCTTGGCCCTACCTTGAATGTTAGTACTGTGTGCTAGCA 364

|||||

48 GlyAsnArgArgPheProValArgGlnProIleLeuGlnTyrArgTr 64

|||||

365 GGAACCCGTAGCGGTTCCGCGTTAGCAGCCCATCTCTCAGTATAGATG 414

|||||

64 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGlu 81

|||||

415 GGACATAATGATAGGCTTGGAGAGCCACAGGCAAGGATGAGAGAGGAGA 464

|||||

81 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 97

|||||

465 ATATGGAAGGATTGGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGG 514

|||||

98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 114

|||||
515 GAAAGCAGTTGAGTCATAGTCTGGGCGACTCAGCACTGATCCCTCA 564
|||||
114 sHisAspHisHisAspGluPheCys.LeuMetPro 125
|||||
565 CCATGACCATCAGCATGAGTTTGCCCTTATGCC 599
|||||

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.AAC03880

seq_documentation_block:

ID AAC03880 standard; cDNA; 662 BP.

AC AAC03880;

DT 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 3878.

DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

PD 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX P-PSDB; AAG03874.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 3878; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from

CC mRNAs encoding secreted proteins. An ORF has been identified within the

CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs

CC derived from 30 different tissues. EST sequences usually correspond

CC mainly to the 3' untranslated region (UTR) of the mRNA because they are

CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not

CC well suited for isolating cDNA sequences derived from the 5' ends of

CC mRNAs and even in those cases where longer cDNA sequences have been

CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from

CC mRNAs with intact 5' ends and can therefore be used to obtain full length

CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,

CC gene therapy and chromosome mapping procedures. They are used to obtain

CC upstream regulatory sequences and to design expression and secretion

XX vectors.

XX Sequence 662 BP; 177 A; 144 C; 192 G; 137 T; 12 other;

alignment_scores:

Quality: 585.50 Length: 128

Ratio: 4.920 Gaps: 1

Percent Similarity: 92.969 Percent identity: 86.719

alignment_block:

US-09-327-750D-32 x AAC03880 ..

Align seg 1/1 to: AAC03880 from: 1 to: 662

1 MetGluSerLysGluLysArgAlaValAsnSerLeuMetGluAsnAl 17

|||||
207 ATGAGTCCAAAGAGGACGCGTTAAACATCTCATCGTGGAAATGT 256
|||||
17 asnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31
|||||
257 CAACACAGGAAATGATGAAAAAGATGAAAGGAGCAGTGCCTAATAAG 306
|||||
31 LysGluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47
|||||
307 GGGAGCCCTTGCCCTACCTTGTGATGTTGTGTAATACTGTGCTGCTAGA 356
|||||
48 GlyAsnArgArgPheProValArgGlnProIleLeuGlnTyrArgTr 64
|||||
357 GGAATCGTAGCGGTTCGCGCTTAGGAGCCCATCTCTGCAGTAGATG 406
|||||
64 PaspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluAla 81
|||||
407 GGATATGATGATAGCTTGGAGACACGAGGAGGATGAGNAGAAGA 456
|||||
81 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 97
|||||
457 ATATGGAAGGATTGGGAGGAGGTGAGACAGCTGATGGAAGAAGCTGAGG 506
|||||
98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 114
|||||
507 GAAAAGCAGTTGATGATCTCTGGGGCGATCAGCACTGACCCCTCA 556
|||||
114 sHisAspHisHisAspGluPheCysLeuMetPro 125
|||||
557 CCATGACCATCAGCATGATGAGTTTTCNNWATGCC 590
|||||

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.AAC03879

seq_documentation_block:

ID AAC03879 standard; cDNA; 698 BP.

AC AAC03879;

DT 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 3877.

DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX P-PSDB; AAG03873.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 3877; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from

CC mRNAs encoding secreted proteins. An ORF has been identified within the

CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs

CC derived from 30 different tissues. EST sequences usually correspond

CC mainly to the 3' untranslated region (UTR) of the mRNA because they are

CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not

CC well suited for isolating cDNA sequences derived from the 5' ends of

CC mRNAs and even in those cases where longer cDNA sequences have been

CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from

CC mRNAs with intact 5' ends and can therefore be used to obtain full length

CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,

CC gene therapy and chromosome mapping procedures. They are used to obtain

CC upstream regulatory sequences and to design expression and secretion

XX vectors.

CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.
 XX
 XX Sequence 698 BP; 187 A; 151 C; 210 G; 139 T; 11 other;

alignment_scores:
 Quality: 585.50 Length: 128
 Ratio: 4.920 Gaps: 1
 Percent Similarity: 92.969 Percent Identity: 86.719

alignment_block:

US-09-327-750D-32 x AAC03879

Align seg 1/1 to: AAC03879 from: 1 to: 698

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
 243 ATGGAGTCCAAAGAGGACGAGCGTTAAACAATCTCATCTGCGAAATCT 292
 17 asanGlnGluasn.....GluGluLysGluGlnValAlaAsnLysG 31
 293 CAACGAGGAAATGATGAAAAAGATGAAAGGACCAAGTGCCTAATAAG 342
 31 LysGluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47
 343 GGGAGCCCTTGGCCCTACCTTGTGATGCTGTGCTGTGCTGTGCTGTG 392
 48 GlyAsnArgArgArgPheProValArgGlnProLeuLeuGlnTyrArgTr 54
 393 GGAATATCGTAGCGGTTCGCGGTAGGACGCCATCTGTCAGTATAGATG 442
 64 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGlu 81
 443 GGATATGATGATAGCTTGGAGAACACACAGGCAAGGATGANAGAAGAGA 492
 81 snMetClnArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 97
 493 ATATGGAAGGATTGGGAGGAGGTGAGACAGCTGTGGAAGCTGAGG 542
 98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 114
 543 GAAAGACAGTTCAGTATAGTCTGCGGCGAGTCAGCACTGACCCCTCA 592
 114 sHisAspHisHisAspGluPheCysLeuMetPro 125
 593 CCATGACCATCATGATGATGTTTCNNWATGCC 626

seq_name: /SID2/ycgdata/geneseq/geneseq/NA2000.DAT:AAC10889

seq_documentation_block:

ID AAC10889 standard; cDNA; 692 BP.

XX AAC10889;

AC AAC10889;

XX 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 14964.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX Homo sapiens.

OS Homo sapiens.

XX EP1033401-A2.

PN EP1033401-A2.

XX

PD 06-SEP-2000.
 XX 21-FEB-2000; 2000EP-0200610.
 XX 26-FEB-1999; 99US-0122487.
 PR (GEST) GENSET.
 PA Dumas Milne Edwards J, Duclert A, Giordano J;
 XX WPI; 2000-500381/45.
 DR New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX Claim 1: SEQ ID 14964; 71pp + CD-ROM; English.
 PS The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 XX
 XX Sequence 692 BP; 184 A; 151 C; 199 G; 146 T; 12 other;

alignment_scores:

Quality: 562.50 Length: 125
 Ratio: 4.849 Gaps: 1
 Percent Similarity: 92.800 Percent Identity: 84.800

alignment_block:

US-09-327-750D-32 x AAC10889

Align seg 1/1 to: AAC10889 from: 1 to: 692

4 LysGluLysArgAlaValAsnSerLeuSerMetGluAsnAlaAsnGlnG 20
 246 CAAAGAGACGAGCGTTAAACAATCTCATCTGCGAAATGTCAACACGAG 295
 20 uasn.....GluGluLysGluGlnValAlaAsnLysGlyGluPro 34
 296 AAATGATGAAAGATGAAAGGACCAAGTGCCTAATAAGGGGAGCCCT 345
 34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50
 346 TGCCCTACCTTTGATGTGTGTAATCTGTGCTGTAGAGGAAATCGT 395
 51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 67
 396 AGCGGCTCCGCTTAGGACGCCCTCTGCTGATATAGTGGATATGAT 445
 67 tHisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGlu 84
 446 GCATAGGCTTGGAGAACCAACAGGCAAGGATGANAGAAGAGAAATATGAAA 495
 84 rGleGlyGluGlnValArgGlnLeuMetGluLysLeuArgGluLysGln 100
 496 GGATTTGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGGGAACGAG 545
 101 LeuSerHisSerLeuArgAlaValSerThrAspProProHisHisAspH 117
 546 TTGAGTATAGTCTGCGGCGAGCTCAGCAGTACGCCCTCCCTCACCATGACCA 595

117 shisaspGluPheCysLeuMetPro 125
 |||||
 596 TCATGATGAGTTTGCNNWATGCC 620

seq_name: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT: AAC06183

seq_documentation_block:

ID AAC06183 standard; cDNA; 421 BP.

AC AAC06183;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 10258.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 XX gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 10258; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

XX Sequence 421 BP; 110 A; 103 C; 124 G; 82 T; 2 other;

alignment_scores:

Quality: 207.50 Length: 72
 Ratio: 3.402 Gaps: 2
 Percent Similarity: 84.722 Percent Identity: 65.278

alignment_block:

US-09-327-750D-32 x AAC06183

Align seg 1/1 to: AAC06183 from: 1 to: 421

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17

207 ATGGAGTCCAAAGAGGAACGAGCGTTAAACAATCTCATCGTGGAAATGT 256

17 aAsnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31
 :|||||
 257 CAACACGAGAAATGATGAAGAAGATGAAGAAGACAGAGTTGCTAATAAG 306
 |||||
 31 LylGluPProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47
 |||||
 307 GGGAGCCCTTGGCCCTACCTTTGAATGTTAGTAATCTGTGTGCTTAGA 356
 |||||
 48 GlyAsnArgArgArgPheProValArgGlnProIleLeuGln.TyrArgT 64
 |||||
 357 GGAACCCGTAG.CGTTCCGGCTTAGGCACACAGTATTAGAGCGACGCT 405
 |||||
 64 rPaspIleMethis 68
 :|||
 406 GCCCAGTGCACAT 419

seq_name: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT: AAC06180

seq_documentation_block:

ID AAC06180 standard; cDNA; 457 BP.

XX AAC06180;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 10255.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 10255; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

XX Sequence 457 BP; 120 A; 110 C; 142 G; 84 T; 1 other;

alignment_scores:

Quality: 207.50 Length: 72
 Ratio: 3.402 Gaps: 2

Percent Similarity: 84.722 Percent Identity: 65.278

alignment_block:

US-09-327-750D-32 x AAC06180

Align seg 1/1 to: AAC06180 from: 1 to: 457

```

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||
243 ATGGAGTCCAAAGAGGACGAGCGTTAAACAATCTCATCGTGGAAAATGT 292
|||||
17 aasnGluGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31
:|||||
293 CAACCCAGGAATAATGATGAAAAAGATGAAAGAGGACGAAAGTTCATAAAG 342
|||||
31 lyGluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47
|||||
343 GGGAGCCCTTGGCCCTACCTTGAATGTTAGTGAATACTGTGTGCTAGA 392
|||||
48 GlyAsnArgArgArgPheProValArgGlnProIleLeuGln.TyrArgT 64
|||||
393 GGAACACCGTAG.CGGTTCGCGTGTAGGCACACAGTATTAGAGGCACCGCT 441
|||||
64 rpaSpilleMethis 68
:|
442 GCCCAGTGACACAT 455

```

seq_name: /STD2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH03517

seq_documentation_block:

ID AAH03517 standard; cDNA: 865 BP.

AC AAH03517;

DT 26-JUN-2001 (first entry)

DE Human cDNA clone (5'-primer) SEQ ID NO:352.

XX Human: primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

PN EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

XX full-length cDNAs defined in the specification, and for the detection

XX and/or diagnosis of the abnormality of the proteins encoded by the

XX full-length cDNAs.

XX Claim 1; SEQ ID 352; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

XX full-length cDNAs defined in the specification. Where a primer set

XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 XX Sequence 865 BP; 239 A; 177 C; 224 G; 221 T; 4 other;

alignment_scores:

Quality: 202.50 Length: 126
 Ratio: 2.382 Gaps: 3
 Percent Similarity: 67.460 Percent Identity: 38.889

alignment_block:

US-09-327-750D-32 x AAH03517

Align seg 1/1 to: AAH03517 from: 1 to: 865

```

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||
191 ATGGAGTCCAAAGAGGACGAGCGTTAAACAATCTCAACGGGGAATATC 240
|||||
17 aasnGluGluAsnGluLysGluGlnValAlaAsnLysGlyGluProL 34
:|||||
241 CCAACAAGAAACAGAGAGGAGGAGCGCCGCCAGCAGAGTGAAGAAG 290
:|
34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50
:|
291 AATCCCGCCATTGGAGGGGGTGAAGCCAGAGCGCTGGAGGAAATATC 340
|||||
51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTyrPaspileMe 67
|||||
341 AGCGCGGGCGAGTAGCGGACTTGTCCCTAATTTTCGATGGCCCATACC 390
|||||
67 thisArg...LeuGlyGluProGlnAlaArgMetArgGluGluAsnMetG 83
:|
391 TAATAGGCATATTGAGCACAATGAAGCGAGA.....GATGATGTAG 431
|||||
83 luArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLys 99
|||||
432 AAAGGTTTGTAGGCGCAGATGATGGAATCAAGAGAAGAGACTAGGGAACAG 481
|||||
100 GlnLeuSerHisSerLeuArgAlaValSerThrAspProHisHisAs 116
:|
482 CAGATGAGCGCACTATATGCGCTTCCAAACCTCTGAACCT.....GA 522
|||||
116 phisHisaspGluPheCysLeuMetPro 125.
:|
523 CAACCATTTATGACTTTGCGCTCATACCT 550

```

seq_name: /STD2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH13750

seq_documentation_block:

ID AAH13750 standard; cDNA: 1229 BP.

XX AC AAH13750;

XX 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:10656.

Percent Similarity: 62.037 Percent Identity: 44.444
alignment_block:
US-09-327-750D-32 x AAF21748
Align seg 1/1 to: AAF21748 from: 1 to: 917
19 GlnGluAsnGluGluLysGluInValAlaAsnLysGlyGluProLeuAl 35
|||||
306 CAGGAAACGAGAGATGGAGCCCTATCGAGATGGAGAGAA...GA 352
35 aLeuProLeuAspAlaGlyGlyTyrCysValProArGlyAsnArgArgA 52
|||||
353 CGCCCTTTGGGAGGAGTGAAGGCCACCGCTGCAGGAAATCGACGG 402
52 rgPheProValArgGlnProIleLeuGlnTyrArgTyrPaspileMetHis 68
|||||
403 GA...CAGGCTGGCGGACTTGGCCCTAATTTTCATGGGCCATACCCAT 449
69 ArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluArgI 85
|||||
450 AGGCAGATCAATGATGGGATGGGTGGAGATGGAGATGATATGGAATATT 499
85 eclyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGlnLeu 102
|||||
500 CATGGAGGAGATGAGAGAAATCAGAGAAACTTAGGAGCTCCAGTTGA 549
102 erHisSerLeuArgAlaValSerThrAsp...ProProHisHisAspHis 117
|||||
550 GGAATTGTCTGCTATCCTTATGGGGAGGCTCTCTAATCACCACCAT 599
118 HisAspGluPheCysLeuMetPro 125
|||||
600 CATGATGAATTTTGCCTTATGCCT 623

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAF21748
seq_documentation_block:
ID AAF21748 standard: DNA; 917 BP.
AC AAF21748;
DT 27-MAR-2001 (first entry)
XX Human breast and ovarian cancer associated antigen gene SEQ ID 135.
DE
XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; antitumor; vulnary; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW Addison's disease; allergy; autoimmune haemolytic anaemia;
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW cardiovascular disorder; wound healing; neurological disease; ds.
XX
OS Homo sapiens.
XX
XX WO200055173-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05881.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI: 2000-611515/58.
XX P-PSDB: AAB58845.
XX
XX New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases.
XX
XX Claim 1; Page 581-582; 1299pp; English.
XX
XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC associated with breast and ovarian cancer. Included in the invention are
CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC isolation and characterisation of the DNA and protein sequences of the
CC invention. The breast and ovarian cancer associated DNA, protein, agonist
CC or antagonist sequences exhibit cytostatic; immunosuppressive;
CC neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC antidiabetic; antiinflammatory; antitumor; vulnary; anticonvulsant;
CC antibacterial; antifungal; antiparasitic and cardiant activity. The
CC polynucleotide and protein sequences are used in the diagnosis of cancer,
CC particularly breast and ovarian cancer. The nucleic acid sequences,
CC proteins, agonists and antagonists may also be used in the diagnosis,
CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemia; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases.
XX
XX Sequence 917 BP; 228 A; 203 C; 270 G; 211 T; 5 other;

alignment_scores:
Quality: 185.50 Length: 108
Ratio: 2.769 Gaps: 3


```

153 AGCGGTTCCCGTTAGCAGCCCATCTCCAGCTATAGATGGGATATGAT 202
67 tHisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 84
|||||
203 GCATAGGCTTGGAGAACACAGGCAAGGATGAGAGAGAGATATGAAA 252
84 rgileGlyGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100
|||||
253 GGATTGGGAGGAGGTGAGACAGTGTGGAAGCTGAGGGAAGAGCAG 302
101 LeuSerHisSerLeuArgAlaValSerThrAspProHisAspH1 117
|||||
303 TTGAGTCTAGTCTGCGGCGAGTCAGCACTGACCCCTCACCATGACCA 352
117 sHisAspGluPheCysLeuMetPro 125
|||||
353 TCATGATGAGTTTTCCTTATGCC 377

```

seq_name: gb_pr:AF220189

seq_documentation_block: 828 bp mRNA PRI 04-MAY-2000
LOCUS AF220189
DEFINITION Homo sapiens uncharacterized hypothalamus protein HBEX2 mRNA, complete cds.

ACCESSION AF220189
VERSION AF220189.1 GI:7689028

KEYWORDS human.
SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Xiao.H., Song.H., Gao.G., Ren.S., Chen.Z. and Han.Z.

JOURNAL A novel gene expressed in human hypothalamus
TITLE Unpublished

REFERENCE 2 (bases 1 to 828)

AUTHORS Xiao.H., Song.H., Gao.G., Ren.S., Chen.Z. and Han.Z.
TITLE Direct Submission

JOURNAL Submitted (30-DEC-1999) Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
FEATURES Location/Qualifiers

source 1..828
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="hypothalamus"
167..544
/codon_start=1
/product="uncharacterized hypothalamus protein HBEX2"
/protein_id="AAF67654.1"
/db_xref="GI:7689028"
/translation="MESKEKRAVNSLSMENANQENEEKQVANKGEPLALPLDAGEYC
VPRGNRRFRVQPQTIQYRNDMHLRGPQARMENMERIGEVRQLMEKLREKQLS
HSLRAVSTDPHHDHDECLMP"
BASE COUNT 256 a 170 c 221 g 181 t
ORIGIN

alignment_scores:
Quality: 653.00 Length: 125
Ratio: 5.266 Gaps: 0
Percent Similarity: 99.200 Percent Identity: 98.400

alignment_block:

US-09-327-750D-32 x AF220189

Align seg 1/1 to: AF220189 from: 1 to: 828

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17

|||||

167 ATGGAGTCCAAGAGAACGAGCAGTAACAGTCTCAGCATGGAATATC 216

|||||

17 aAsnGlnGluAsnGluGluLysGluGlnValAlaAsnLysGlyGluProL 34

```

|||||
217 CAACCAAGAAAATGAAGAAAAGGAGCAAGTTGCTAATAAAGGGAGCCCT 266
34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50
|||||
267 TGGCCCTCCCTTTGGATGCTGTGAATCTGTGTGCTAGAGGAATCGT 316
51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 67
|||||
317 AGGCGGTTCCCGTTAGCAGCCCATCTCCAGTATAGATGGGATATGAT 366
67 tHisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 84
|||||
367 GCATAGGCTTGGAGAACACAGGCAAGGATGAGAGAGAGATATGAAA 416
84 rgileGlyGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100
|||||
417 GGATTGGGAGGAGGTGAGACAGTGTGGAAGCTGAGGGAAGAGCAG 466
101 LeuSerHisSerLeuArgAlaValSerThrAspProHisAspH1 117
|||||
467 TTGAGTCTAGTCTGCGGCGAGTCAGCACTGACCCCTCACCATGACCA 516
117 sHisAspGluPheCysLeuMetPro 125
|||||
517 TCATGATGAGTTTTCCTTATGCC 541

```

seq_name: gb_pr:HS198P4

seq_documentation_block:

LOCUS HS198P4 35714 bp DNA PRI 16-DEC-2000
DEFINITION Human DNA sequence from clone RP1-198P4 on chromosome Xq22 Contains a CpG Island, complete sequence.

ACCESSION AL008708

VERSION AL008708.4 GI:4826445

KEYWORDS HIG; CpG Island.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Corby.N.

TITLE Direct Submission

JOURNAL Submitted (21-JAN-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clones@sanger.ac.uk

On May 13, 1999 this sequence version replaced GI:4582117.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep/ RP1-198P4 is from the library RCI-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see
VECTOR: pcypac2
http://pacpac.med.buffalo.edu/

IMPORTANT: This sequence is not the entire insert of clone

RPI-198P4 It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
The true left end of clone LLOXNC01-221F2 is at 35615 in this sequence. The true right end of clone LLOXNC01-73B8 is at 100 in this sequence.

FEATURES

Source

Location/Qualifiers

1..35714
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q22"
/clone="RPI-198P4"
/clone_lib="RPCI-1"

repeat_region
1..258
/note="LIPB3 repeat: matches 5666..5925 of consensus"
256..537
/note="LIPB3 repeat: matches 5391..5672 of consensus"
1491..1749
/note="AluSc repeat: matches 39..297 of consensus"
2182..3810
/note="LIPB3 repeat: matches 4533..6150 of consensus"
5169..6687
/note="LIM2 repeat: matches 4798..6308 of consensus"
6976..7665
/note="LIM2 repeat: matches 4113..4798 of consensus"
8039..9213
/note="LIM2 repeat: matches 2945..4113 of consensus"
9523..10437
/note="LIM2 repeat: matches 2016..2945 of consensus"
10736..11494
/note="LIM2 repeat: matches 1262..2016 of consensus"
11807..12251
/note="LIM2 repeat: matches 310..755 of consensus"
12259..12331
/note="HERVH21 repeat: matches 5..77 of consensus"
12332..12796
/note="LTR21A repeat: matches 1..490 of consensus"
16162..16371
/note="105 copies 2 mer tt 55% conserved"
17121..17424
/note="MER33 repeat: matches 1..322 of consensus"
18259..18290
/note="16 copies 2 mer ca 87% conserved"
18505..18672
/note="MER63A repeat: matches 1..181 of consensus"
19478..19814
/note="MER2 repeat: matches 1..345 of consensus"
21061..21101
/note="HERVH21 repeat: matches 42..85 of consensus"
21247..21781
/note="MLT1E repeat: matches 44..568 of consensus"
24465..24492
/note="Single clone region"
Single clone region
24531..24562
/note="Weak data"
Weak data
24563
/note="Single clone region"
Single clone region
24711..24865
/note="region covered by sublines from a PAC DNA PCR product only at 12x coverage."
region covered by sublines from a PAC DNA PCR product only at 12x coverage.
24758..25460
/note="Cpc island"
/evidence="not_experimental"
24875..24940
/note="33 copies 2 mer cc 66% conserved"
26795..26926
/note="66 copies 2 mer ct 63% conserved"

repeat_region
27088..27133
/note="23 copies 2 mer tg 100% conserved"
27461..27777
/note="LIMB8 repeat: matches 5851..6171 of consensus"
27802..28232
/note="LIMC/D repeat: matches 5192..5403 of consensus"
28536..28930
/note="LIMC/D repeat: matches 4802..5192 of consensus"
29240..29341
/note="TIGER1 repeat: matches 484..587 of consensus"
29603..30740
/note="TIGER1 repeat: matches 587..1757 of consensus"
31027..31365
/note="TIGER1 repeat: matches 1757..2105 of consensus"
31664..31691
/note="TIGER1 repeat: matches 2105..2132 of consensus"
31809..32015
/note="TIGER1 repeat: matches 2132..2246 of consensus"
32323..32389
/note="TIGER1 repeat: matches 2246..2313 of consensus"
34353..34636
/note="MLT1-INTERNAL repeat: matches 607..889 of consensus"
35149..35376
/note="MLT1-INTERNAL repeat: matches 1001..1233 of consensus"
35523..35670
/note="MLT1C repeat: matches 8..159 of consensus"
BASE COUNT 9708 a 7094 c 6805 g 12107 t
ORIGIN

alignment_scores
Quality: 653.00 Length: 125
Ratio: 5.266 Gaps: 0
Percent Similarity: 99.200 Percent Identity: 98.400

alignment_block:
US-09-327-750D-32 x HS198P4/rev ..
Align seg 1/1 to reverse of: HS198P4 from: 1 to: 35714

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||
24358 ATGGAGTCCAAAGAGAAAGAGCAGTAAACAGTCTCAGCATGGAAATGC 24309
|||||
17 aAsnGlnGluAsnGluGluLysGluGlnValAlaAsnLysGlyGluProL 34
|||||
24308 CAACCAACAANAATGAGAAAGAGGACAGTCTCTAATAAGGGAGCCCT 24259
|||||
34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50
|||||
24258 TGGCCCTCCCTTGGATGCTGTGTAATCTGTGCTAGAGAAATCGT 24209
|||||
51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTTPAspIleMe 67
|||||
24208 AGCGGTTCGCCCTAGGAGCCCATCTTCACATATAGTGGATATGAT 24159
|||||
67 tHisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 84
|||||
24158 GCATAGGCTTGGAGAACCCAGCAGGAGATGAGAGAGAGAAATATGAAA 24109
|||||
84 rgleGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100
|||||
24108 GGATGGGAGAGAGTGACACAGCTGTATGGAAAGCTGAGGGGAAAGCAG 24059
|||||
101 LeuSerHisSerLeuArgAlaValSerThrAspProHisAspHis 117
|||||
24058 TTGAGTCATAGTCTGGGGCAGTCAGCACTGACCCCTCACCATTGACCA 24009
|||||
117 sHisAspGluPheCysLeuMetPro 125
|||||
24008 TCATGATGAGTTTTCCTTATGCC 23984

```
seq_name: gb_pr:AF237783
seq_documentation_block: 791 bp mRNA PRI 02-SEP-2000
LOCUS AF237783 Homo sapiens brain-expressed protein BEX1 (BEX1) mRNA, complete cds.
DEFINITION Homo sapiens brain-expressed protein BEX1 (BEX1) mRNA, complete cds.
ACCESSION AF237783
VERSION AF237783.1 GI:9963898
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 791)
AUTHORS Yang,Q.S., Ying,K., Xie,Y. and Mao,Y.M.
TITLE A Novel Human X-linked Brain Expressed Protein
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 791)
AUTHORS Mao,Y.M., Xie,Y., Yang,Q.S., Wu,H., Lin,S. and Ying,K.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2000) Genetic Research Institute, Fudan University, 220 Handan Rd., Shanghai 200433, P.R.China
FEATURES
source
1..791
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="between DXS990 and DX1059 by RH (Stanford G3)"
1..791
/gene="BEX1"
172..549
/gene="BEX1"
/codon_start=1
/product="brain-expressed protein BEX1"
/db_xref="AG09752.1"
/db_xref="GI:9963898"
/translation="MESKEKLVNSLSENAQNEKEQVANKGEPLALPLDAGSYC
VPGNRRFRVROPILQYRWDMHRLGEPQARMRENERIGEVRLMEKLEKLS
HSLRAVSTDPHDDHDFCLMP"
BASE COUNT 214 a 172 c 218 g 187 t
ORIGIN

alignment_scores:
Quality: 639.00 Length: 125
Ratio: 5.238 Gaps: 0
Percent Similarity: 97.600 Percent Identity: 96.800
alignment_block:
US-09-327-750D-32 x AF237783 ..
Align seg 1/1 to: AF237783 from: 1 to: 791
1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||
172 ATGGAGTCCAAAGAGAACTAGCAGTAAACAGTCTCAGCATGGAATGC 221
|||||
17 aAsnGlnGluAsnGluGluLysGluGlnValAlaAsnLysGlyGluProL 34
|||||
222 CAACCAAGAAATGAAGAAAGGACCAAGTTGCTAATAAAGGGAGCCCT 271
|||||
34 euAlaLeuProLeuAspAlaGlyGlyTyrCysValProArgGlyAsnArg 50
|||||
272 TGGCCCTCCCTTTGATGCTGCTGAATCTGTGTGCTAGAGGAAATCGT 321
|||||
51 ArgArgPheProValArgLysProLleLeuGlnTyrArgTyrAspIleMe 67
|||||
322 AGCGGTTCCGCGTTAGCAGCCCATCTCCAGTATAGATGGGATATGAT 371
|||||
67 tHisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGlu 84
|||||
372 GCATAGGCTTGAGAACACAGGCAAGGATGAGAGAGAGAAATATGNA 421
|||||
```

```
84 rgIleGlyGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100
|||||
422 GGATTGGGAGGGGTGAGACAGCTGATGGAAGAGCTGAGGAAACGAG 471
|||||
101 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspH1 117
|||||
472 TTGAGTCATAGTCTCGGGCAGTCAGCACTGACCCCTCACCATGACCA 521
|||||
117 sHisAspGluPheCysLeuMetPro 125
|||||
522 TCATGATGAGTTTGCCTTATGCC 546
|||||
seq_name: gb_pr:AF251053
seq_documentation_block:
LOCUS AF251053 744 bp mRNA PRI 15-APR-2001
DEFINITION Homo sapiens X-linked protein mRNA, complete cds.
ACCESSION AF251053
VERSION AF251053.1 GI:13625167
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 744)
AUTHORS Mao,Y., Xie,Y., Zhou,Z., Zhao,W., Zhao,S., Wang,W., Huang,Y.,
Wang,S., Tang,R., Chen,X. and Wu,C.
TITLE Direct Submission
JOURNAL Submitted (29-MAR-2000) Institute of Genetics, School of Life Sciences, Fudan University, 220 Handan Road, Shanghai 200433, P.R. China
FEATURES
source
1..744
/organism="Homo sapiens"
/db_xref="taxon:9606"
101..487
/codon_start=1
/product="X-linked protein"
/protein_id="AAK34943.1"
/db_xref="GI:13625168"
/translation="MESKEERALNNLVENQENDEKDEQVANKGEPLALPLNYS
EYCVPRGNRRFRVROPILQYRWDMHRLGEPQARMRENERIGEVRLMEKLEK
QLSHSLRAVSTDPHDDHDFCLMP"
BASE COUNT 222 a 134 c 208 g 180 t
ORIGIN

alignment_scores:
Quality: 602.50 Length: 128
Ratio: 4.898 Gaps: 1
Percent Similarity: 96.094 Percent Identity: 89.062
alignment_block:
US-09-327-750D-32 x AF251053 ..
Align seg 1/1 to: AF251053 from: 1 to: 744
1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||
101 ATGGAGTCCAAAGAGAGAGAGCGTAAACAATCTCATCGTGGAAATGT 150
|||||
17 aAsnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31
|||||
151 CAACCAAGGAAATGATGAAAAAGAGTGAAGAGAGCAAGTTGCTTAATAAG 200
|||||
31 lyGluProLeuAlaLeuProLeuAspAlaGlyGlyTyrCysValProArg 47
|||||
201 GGGAGCCCTTGGCCCTACCTTTGATGTTAGTGAATACTGTGCTCCTAGA 250
|||||
48 GlyAsnArgArgPheProValArgGlnProLleLeuGlnTyrArgTyr 64
|||||
251 GGAACCGCTAGCGGTTCCGGCTTAGCAGCCCATCTCTGCTAGTATAGATG 300
|||||
```

```

64 pspilleMethHsArgLeuGluGluProGlnAlaAaMetArgGluGluA 81
301 GGACATAATGCTAGGCTGGAGAGCCACAGGCAAGATGAGAGAGGAGA 350
81 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 97
351 ATATGAAAGGATTGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGG 400
98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProH1 114
401 GAAAGCAGCTTGAAGTCTAGTCTCGGCGCAGTGCAGCTGATCCCTCA 450
114 shIsAspHisAspGluPheCysLeuMetPro 125
451 CCATGACCATCAGATGAGTTTGCCTTATGCC 484

seq_name: gb_pat:AX078272
seq_documentation_block:
LOCUS AX078272 898 bp DNA PAT 22-FEB-2001
DEFINITION Sequence 76 from Patent WO0107471.
ACCESSION AX078272
VERSION AX078272.1 GI:13157963
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Hillman,J.L., Lal,P., Tang,Y.T., Yue,H., Au-Young,J., Bandman,O.,
Arimai,Y., Yang,J., Lu,D.A., Baughn,M.R., Patterson,C. and Shah,P.
TITLE Cell cycle and proliferation proteins
JOURNAL Patent: WO 0107471-A 76 01-FEB-2001;
Incyte Genomics, Inc. (US)
FEATURES
source
1..898
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 5664154CB1"
BASE COUNT 250 a 186 c 251 g 211 t
ORIGIN

```

```

alignment_scores:
Quality: 602.50 Length: 128
Ratio: 4.898 Gaps: 1
Percent Similarity: 96.094 Percent Identity: 89.062

alignment_block:
us-09-327-750d-32 x AX078272 ..
Align seg 1/1 to: AX078272 from: 1 to: 898

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
254 ATGACTCCAAAGAGAACGAGCGTAAACAACTCATCTCGTGGAAATGT 303
17 AsnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31
304 CAACAGGAAATATGATGAAAAAGATGAAAGGAGCAAGTTGCTAATAAG 353
31 llyCluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47
354 GGAGGCCCTGGCCCTACTTGTGAATGTAGTACTGTGTCCTAGG 403
48 GlyAsnArgArgPheProValArgGlnProLeuGlnTyrArgTr 64
404 GGAACCGTAGGCGGTTCGCGCTAGGAGCCCATCTCGTGCAGTAGATG 453
64 pspilleMethHsArgLeuGluGluProGlnAlaAaMetArgGluGluA 81
454 GGACATAATGCTAGGCTGGAGAGCCACAGGCAAGATGAGAGAGGAGA 503

```

```

81 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 97
504 ATATGAAAGGATTGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGG 553
98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProH1 114
554 GAAAGCAGCTTGAAGTCTAGTCTCGGCGCAGTGCAGCTGATCCCTCA 603
114 shIsAspHisAspGluPheCysLeuMetPro 125
604 CCATGACCATCAGATGAGTTTGCCTTATGCC 637

seq_name: gb_pr:HSV870H8
seq_documentation_block:
LOCUS HSV870H8 31321 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from cosmid V870H8, between markers DXS366 and
DXS87 on chromosome X contains ESTs.
ACCESSION 270233
VERSION 270233.1 GI:1235542
KEYWORDS X.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 31321)
Whiteley,M.
TITLE Direct Submission
JOURNAL
Submitted (19-MAR-1995) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT
IMPORTANT: This sequence is not the entire insert of clone V870H8.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone V870H8 is at 1 in this sequence. The
true left end of clone V693A8 is at 31221.
V870H8 is from the human chromosome X-specific cosmid library.
FEATURES
Location/Qualifiers
1..31321
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone="GHC-870H8"
/clone_lib="SCCV"
475..701
/note="L1 element fragment"
repeat_region
802..988
/note="L1 element fragment"
repeat_region
2533..4138
/note="match: multiple ESTs"
misc_feature
4189..4246
/note="29 copies of 2 mer 91 % conserved"
repeat_region
7496..7874
/note="L1 element fragment"
repeat_region
8022..8108
/note="L1 element fragment"
repeat_region
8302..8493
/note="L1 element fragment"
repeat_region
8585..8671
/note="L1 element fragment"
repeat_region
8769..9010
/note="L1 element fragment"
repeat_region
9259..9384
/note="L1 element fragment"
repeat_region
9624..9731
/note="MLTIC element fragment"
repeat_region
9781..9948
/note="MLTIC element fragment"
repeat_region
9890..9948
/note="MLT1B element fragment"
repeat_region
10418..10592

```

```

/partial
/note="Alu repeat: matches 298. .108 of consensus"
10593. .10688
/partial
/note="Alu repeat: matches 96. .1 of consensus"
10698. .10781
/note="L1 element fragment"
11112. .11403
/note="Alu repeat: matches 1. .308 of consensus"
11922. .12004
/note="L1 element fragment"
12062. .12308
/note="L1 element fragment"
12423. .12707
/partial
/note="Alu repeat: matches 302. .1 of consensus"
12710. .12916
/note="L1 element fragment"
12979. .13081
/note="L1 element fragment"
13473. .13532
/note="MT2A1 element fragment"
13871. .13928
/note="L1 element fragment"
18180. .18417
/note="L1 element fragment"
18778. .19056
/note="Alu repeat: matches 1. .308 of consensus"
22279. .22350
/note="L1 element fragment"
22449. .22995
/note="L1 element fragment"
23239. .23374
/note="L1 element fragment"
23371. .23606
/note="MT1B element fragment"
23395. .23541
/note="MT1A element fragment"
23705. .23778
/note="MT1A element fragment"
23805. .23960
/note="L1 element fragment"
24011. .24541
/note="L1 element fragment"
24638. .24770
/partial
/note="Alu repeat: matches 150. .1 of consensus"
24829. .30975
/note="L1 element fragment"
BASE COUNT 9462 a 6147 c 6129 g 9583 t
ORIGIN

alignment_scores:
  Quality: 602.50      Length: 128
  Ratio: 4.898        Gaps: 1
  Percent Similarity: 96.094      Percent Identity: 89.062

alignment_block:
US-09-327-750D-32 x HSV870H8
Align seg 1/1 to: HSV870H8 from: 1 to: 31321

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||
3508 ATGGAGTCAAGAGGACGAGCTTAACAATCTCATGCTGGAAAATCT 3557
|||||
17 asnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31
|||||
3558 CAACGAGGAATGATGAAAGAGATGAAAGGACGAGCTGCTATATAAG 3607
|||||
31 lyGluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47
|||||

```

```

3608 GGGAGCCCTTGCCCTACCTTTGAATGTTAGTGAATACTGTGTCCTAGA 3657
48 GLVAsnAlaGArgAlaGlnProValArgGlnProIleLeuGlnTyrArgTr 64
|||||
3658 GGAACCCGTAGCGGTTCCCGGTTAGGACGCCATCTCGCAGTATAGATG 3707
64 pasPleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 81
|||||
3708 GGACATAATGTCATAGCTTGAGAGGCCACAGGCAAGGATGAGAGAGAGA 3757
81 snMetGluArgIleGlyGluGlnValArgGlnLeuMetGluLysLeuArg 97
|||||
3758 ATATGAAAGGATTGGGAGGAGCTGACACAGCTGATGGAAAGCTGAGG 3807
98 GluLysGluLeuSerHisSerLeuArgAlaValSerThrAspProH 114
|||||
3808 GAAAGAGCATTTAGTCATAGTCTCGGGCAGTCAGCAGCTGATCCCTCA 3857
114 shisAspHisHisAspGluPheCysLeuMetPro 125
|||||
3858 CCATGACCATCAGATGAGTTTGCCTTATGCC 3891

seq_name: gb_pr:AL133348
seq_documentation_block:
LOCUS AL133348 40584 bp DNA PRI 06-MAR-2000
DEFINITION Human DNA sequence from clone RPL-79p11 on chromosome Xq21.32-22.1.
Contains the gene for a novel protein similar to mouse Bex2
(brain-expressed X-linked protein 2), ESTs, STSS, GSSs and a
putative CpG island, complete sequence.
ACCESSION AL133348
VERSION AL133348.8 GI:7076386
KEYWORDS HTG; Bex2; CpG island.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 40584)
AUTHORS Wilson, S.
TITLE Direct Submission
JOURNAL
COMMENT
On Feb 24, 2000 this sequence version replaced gi:6997869.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/Celegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
RPL-79p11 is from the library RPCI-1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone
RPL-79p11 it may be shorter because we only sequence overlapping
sections once, or longer because we arrange for a small overlap
between neighbouring submissions.

```

The true left end of clone LLOXNC01-105G4 is at 40485 in this sequence. The true right end of clone LLOXNC01-177E8 is at 100 in this sequence.

FEATURES

source

Location/Qualifiers

```

1. .40584
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q21.32-22.1"
/clone="RPL-79P11"
/clone_lib="RPC1-1"
19. .292
/notes="AluJo repeat: matches 3. .275 of consensus"
repeat_region
470. .1284
/notes="LIMB5 repeat: matches 5318. .6157 of consensus"
repeat_region
1295. .1898
/notes="LIPAL5 repeat: matches 5556. .6157 of consensus"
repeat_region
1899. .2345
/notes="LIPAL5 repeat: matches 4922. .5370 of consensus"
repeat_region
2323. .2698
/notes="LIMB5 repeat: matches 4960. .5343 of consensus"
repeat_region
2927. .3065
/notes="MER7A repeat: matches 164. .305 of consensus"
repeat_region
3655. .9795
/notes="LIPAL4 repeat: matches 2. .6142 of consensus"
repeat_region
9866. .9982
/notes="FLAM_C repeat: matches 6. .124 of consensus"
repeat_region
9989. .10254
/notes="L1 repeat: matches 4027. .4293 of consensus"
repeat_region
10253. .10852
/notes="LIM4 repeat: matches 4755. .5333 of consensus"
repeat_region
10853. .11260
/notes="MLT1B repeat: matches 3. .390 of consensus"
repeat_region
11261. .11474
/notes="LIM4 repeat: matches 4538. .4755 of consensus"
repeat_region
11480. .11563
/notes="MSTB repeat: matches 1. .80 of consensus"
repeat_region
11564. .12475
/notes="L1 repeat: matches 3039. .3971 of consensus"
misc_feature
14463. .14861
/notes="match: GSS: Em:AQ729215"
repeat_region
15578. .15854
/notes="AluJo repeat: matches 6. .289 of consensus"
misc_feature
complement(16007. .16556)
repeat_region
16197. .16569
/notes="LIM4 repeat: matches 4267. .4627 of consensus"
repeat_region
16629. .16714
/notes="LIM4c repeat: matches 1578. .1666 of consensus"
repeat_region
16902. .17133
/notes="LIM4 repeat: matches 0. .242 of consensus"
repeat_region
17536. .17728
/notes="MIR repeat: matches 1. .191 of consensus"
repeat_region
17746. .18323
/notes="L2 repeat: matches 1599. .2175 of consensus"
repeat_region
18706. .18755
/notes="LIMB3 repeat: matches 6034. .6083 of consensus"
repeat_region
18756. .19236
/notes="MER6B repeat: matches 1. .486 of consensus"
repeat_region
19237. .21108
/notes="MER6B-internal repeat: matches 4808. .6676 of consensus"
repeat_region
21109. .21163
/notes="MLT2B repeat: matches 391. .445 of consensus"
repeat_region
21164. .21555
/notes="MER6B repeat: matches 3. .391 of consensus"
repeat_region
21557. .21662
/notes="LIMB3 repeat: matches 6080. .6185 of consensus"
repeat_region
21665. .21761
/notes="LIM repeat: matches 5299. .5392 of consensus"
repeat_region
21762. .21929
/notes="LIMB8 repeat: matches 5902. .6063 of consensus"
repeat_region
21930. .22212

```

```

/notes="Alusg repeat: matches 3. .284 of consensus"
22213. .22321
/notes="LIMB8 repeat: matches 6063. .6173 of consensus"
22333. .22740
/notes="LIMD2 repeat: matches 5867. .6288 of consensus"
22744. .23094
/notes="LIMD repeat: matches 2. .342 of consensus"
23218. .23524
/notes="Alusg repeat: matches 1. .305 of consensus"
23603. .23947
/notes="LIM4 repeat: matches 4916. .5184 of consensus"
23948. .24218
/notes="AluJo repeat: matches 1. .282 of consensus"
24219. .24510
/notes="LIM4 repeat: matches 4601. .4916 of consensus"
24508. .24621
/notes="LIM4 repeat: matches 4374. .4488 of consensus"
24650. .25077
/notes="MLT1C repeat: matches 1. .466 of consensus"
25078. .25914
/notes="LIM4 repeat: matches 3497. .4346 of consensus"
25920. .26741
/notes="LIM4 repeat: matches 2202. .3065 of consensus"
26762. .26906
/notes="LIPAL7 repeat: matches 6001. .6145 of consensus"
26907. .27135
/notes="LIPAL5 repeat: matches 5674. .5902 of consensus"
27266. .27509
/notes="LIM4 repeat: matches 2164. .2080 of consensus"
27557. .27769
/notes="LIMEC repeat: matches 253. .460 of consensus"
28120. .28167
/notes="L2 repeat: matches 2703. .2750 of consensus"
29043. .29349
/notes="MER5B repeat: matches 1. .341 of consensus"
30389. .30452
/notes="16 copies 4 mer tata 89% conserved"
complement(30502)
polyA_site
/genes="dj79p11.1"
mRNA
complement(join(30502. .31137,31514. .31589,32016. .>32134))
/genes="dj79p11.1"
/notes="match: ESTs: Em:AA317587 Em:AA358632 Em:AI214048
Em:AI29470 Em:AA319216 Em:AA317625 Em:D81392 Em:AA317418
Em:AA366528 Em:C15186 Em:AI291270 Em:AA160660 Em:AA612605
Em:AA160832 Em:AA384405 Em:AA160657 Em:AA081744
Em:AI291126 Em:C15591 Em:AA384268 Em:AI929703 Em:AI302026
Em:W19547 Em:AI141727 Em:AA946933 Em:AI141583 Em:H61106
Em:W60581 Em:AI826617 Em:AA706545 Em:AA64771 Em:AA862327
Em:W68380 Em:AA641652 Em:AA384269 Em:AA527696 Em:W48778
Em:AA633989 Em:AI659486 Em:AW167301 Em:AI929428
Em:AA379411 Em:C15948 Em:AI745669 Em:AW024587 Em:AW379207;
match: proteins: Tr:Q9Y516 Tr:Q9Y517"
/evidence=not_experimental
/product="dj79p11.1 (novel protein similar to mouse Bex2
(brain-expressed x-linked protein 2))"
complement(30502. .32134)
genes
/genes="dj79p11.1"
30507. .30988
/notes="match: STS: Em:G24641; match: STS: Em:T86927"
30508. .30852
/notes="match: STS: Em:R43117; match: STS: Em:G23964"
complement(30518. .30523)
polyA_signal
/genes="dj79p11.1"
30541. .30922
/notes="match: STS: Em:H61107"
31543. .32262
/notes="CpG island"
/evidence=not_experimental
31647. .31730
/notes="42 copies 2 mer cc 64% conserved"
complement(32655. .33316)
misc_feature
/notes="match: GSS: Em:AQ489303"
33027. .34514

```

/note="LIME2 repeat: matches 4631. .6155 of consensus"
 34528. .34786
 /note="LIME3A repeat: matches 4404. .4664 of consensus"
 34787. .35085
 /note="AluSg repeat: matches 1. .300 of consensus"
 35086. .35412

alignment_scores:
 Quality: 602.50 Length: 128
 Ratio: 4.898 Gaps: 1
 Percent Similarity: 96.094 Percent Identity: 89.062

alignment_block:

US-09-327-750D-32 x AL133348/rev ..

Align seg 1/1 to reverse of: AL133348 from: 1 to: 40584

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
 |||||
 31132 ATGGAGTCCAAAGAGGACGAGCGTTAAACAATCTCATCGTGGAAATGT 31083

17 aAsnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31
 |||||

31082 CAACCAAGGAAATGATGAAAGATGAAAGAGGACCAAGTTGCTTAATAAG 31033

31 lYgluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47
 |||||
 31032 GGGAGCCCTGGCCCTACCTTGAATCTTAGTGAATCTGTGTGCCTAGA 30983

48 GlyAsnArgArgPheProValArgGlnProLeuGlnTyrArgTr 64
 |||||

30982 GGAACCCCTAGCGGTTCCGCGTTAGGACGCCCATCTCGCAGTATAGTG 30933

64 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGlu 81
 |||||

30932 GGACATAATGATAGCTGTGGAGGCCACAGGCAAGGATGAGAGAGAGA 30883

81 sMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 97
 |||||

30882 ATATGGAAGGATTGGGAGAGGTGAGCAGCTGATGGAAAGCTGAGG 30833

98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 114
 |||||

30832 GAAAGCAGTTGAGTCATAGTTTCGGGGCAGTCAGCACTATCCCCCTCA 30783

114 sHisAspHisAspGluPheCysLeuMetPro 125

30782 CCATGACCATCAGATGAGTTTGCCTTATGCC 30749

seq_name: gb_hcg:HSU80B1

seq_documentation_block:

LOCUS HSU80B1 41029 bp DNA HTG 10-JUL-2001
 DEFINITION Homo sapiens chromosome X clone LLOXNC01-80B1 map q22.1-22.3, ***
 SEQUENCING IN PROGRESS ***, in ordered pieces.

ACCESSION AL022169

VERSION AL022169.3 GI:13276704

KEYWORDS HTG: HTGS_PHASE2; HTGS_CANCELLED.

SOURCE human

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 41029)

Bird.C.

Direct Submission

Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Mar 12, 2001 this sequence version replaced gi:12750927.

COMMENT

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: CU80B1
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator ABI; 19% of reads
 Consensus quality: 41009 bases at least Q40
 Consensus quality: 41023 bases at least Q30
 Consensus quality: 41027 bases at least Q20
 Insert size: 41029; sum-of-contigs
 Insert size: 51751; 0.6% error; agarose-pp
 Quality coverage: 9.67x in Q20 bases; sum-of-contigs Quality
 coverage: 7.67x in Q20 bases; agarose-pp

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

FEATURES

Source

1. 41029
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="q22.1-22.3"
 /clone="LLOXNC01-80B1"
 /clone_lib="LLOXNC01"

misc_feature

1. 41029
 /note="assembly_fragment:00890"
 BASE COUNT 13249 a 7796 c 7781 g 12203 t
 ORIGIN

alignment_scores:

Quality: 602.50 Length: 128
 Ratio: 4.898 Gaps: 1
 Percent Similarity: 96.094 Percent Identity: 89.062

alignment_block:

US-09-327-750D-32 x HSU80B1 ..

Align seg 1/1 to: HSU80B1 from: 1 to: 41029

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17

18412 ATGGAGTCCAAAGAGGACGAGCGTTAAACAATCTCATCGTGGAAATGT 18461

17 aAsnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31

18462 CAACCAAGGAAATGATGAAAGATGAAAGAGGACCAAGTTGCTTAATAAG 18511

31 lYgluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47

18512 GGGAGCCCTGGCCCTACCTTGAATCTTAGTGAATCTGTGTGCCTAGA 18561

48 GlyAsnArgArgPheProValArgGlnProLeuGlnTyrArgTr 64

18562 GGAACCCCTAGCGGTTCCGCGTTAGGACGCCCATCTCGCAGTATAGTG 18611

64 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGlu 81

18612 GGACATAATGATAGCTGTGGAGGCCACAGGCAAGGATGAGAGAGAGA 18661

81 sMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 97

18662 ATATGGAAGGATTGGGAGAGGTGAGACACCTGATGGAAAGCTGAGG 18711

98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 114

18712 GAAAGCAGTTGAGTCATAGTCTCGGGCAGTCAGCACTATCCCCCTCA 18761

114 sHisAspHisAspGluPheCysLeuMetPro 125

```

18762 CCATGACCATCAGGATGAGTTTCCTTATGCC 18795
seq_name: gb_ro:AF097439
seq_documentation_block:
LOCUS AF097439 785 bp mRNA 13-APR-1999
DEFINITION Mus musculus brain expressed X-linked protein 2 (Bex2) mRNA,
complete cds.
ACCESSION AF097439
VERSION AF097437.1 GI:4580591
KEYWORDS house mouse.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 785)
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Bex1, a gene with increased expression in parthenogenetic embryos,
is a member of a novel gene family on the mouse X chromosome
JOURNAL Hum. Mol. Genet. 8 (4), 611-619 (1999)
MEDLINE 99172070
REMARK Erratum:[published erratum appears in Hum Mol Genet 1999
May;8(5):943]]
REFERENCE 2 (bases 1 to 785)
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1998) Cancer Unit, Queensland Institute of
Medical Research, Herston Rd, Brisbane, Qld 4029, Australia
FEATURES
source
1..785
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="dbEST:W48832"
/chromosome="X"
/map="near Pip"
/dev_stage="embryo; 15.5 dpc"
1..785
/gene="Bex2"
139..528
/gene="Bex2"
/codon_start=1
/product="brain expressed X-linked protein 2"
/protein_id="AAD24430.1"
/db_xref="GI:4580592"
/translation="MESKVEQGVKLNLMNENHOBKEEKEKPODASKRDPVALPREA
GDYVPRGRRFRVRQPIVHRWDLMLQRYGEPQGRMRNENVRFGDVRQLMEKLE
RQLSHSLRAVSTDPDHPHDEFCIMP"
BASE COUNT 212 a 163 c 235 g 175 t
ORIGIN
alignment_scores:
Quality: 496.00 Length: 129
Ratio: 4.168 Gaps: 2
Percent Similarity: 92.248 Percent Identity: 72.093
alignment_block:
US-09-327-750D-32 x AF097439 ..
Align seg 1/1 to: AF097439 from: 1 to: 785
1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
139 ATGGAGTCCAAAAGTGACAGCGCTGAAATCTCAACATGGAGATGA 188
17 aAsnGluGluAsnGluGluLysGlu.....GlnValAlaAsnLysG 31
189 CCATCAGGAAGAGGAGGAAGAGAAAGCAAGCAAGATGCTAGCAAAA 238
31 lyGluProLeu...AlaLeuProLeuAspAlaGlyGluTyrCysValPro 46

```

```

239 GGGATCCGATTGTCGCCCTTCGGAAGCTGGAGACTACTAGTGCCT 288
47 ArgGlyAsnArgArgArgPheProValArgGlnProIleLeuGlnTyrAl 63
289 AGAGGAGGTCCGAGCGGTTCGGGTCGGCAGCCCATCGTGCCTACAG 338
63 gtrPaspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGlu 80
339 ATGGGACCTGATGCATAGGTGGGGAGCCCGAGGAGGATGAGAGAGG 388
80 LuasnMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeu 96
389 AGAACGTACAGAGGTGGGGATGATGTGACACAGCTCATGGAGAAGCTG 438
97 ArgGluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 113
439 AGGGAAGCAGCTGAGCCACAGCCTCGCGGCGGTAGCACTGACCCGCC 488
113 OHisHisAspHisHisAspGluPheCysLeuMetPro 125
489 TCATCATGACCAACCATGATGAGTTTTCCTTATGCC 525
seq_name: gb_ro:AF097437
seq_documentation_block:
LOCUS AF097437 2269 bp DNA 05-AUG-1999
DEFINITION Mus musculus Bex1 protein (Bex1) gene, complete cds.
ACCESSION AF097437
VERSION AF097437.1 GI:5702155
KEYWORDS house mouse.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 2269)
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Bex1, a gene with increased expression in parthenogenetic embryos,
is a member of a novel gene family on the mouse X chromosome
JOURNAL Hum. Mol. Genet. 8 (4), 611-619 (1999)
MEDLINE 99172070
REMARK Erratum:[published erratum appears in Hum Mol Genet 1999
May;8(5):943]]
REFERENCE 2 (bases 1 to 2269)
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1998) Cancer Unit, Queensland Institute of
Medical Research, Herston Rd, Brisbane, Qld 4030, Australia
FEATURES
source
1..2269
/organism="Mus musculus"
/strain="129SVJ"
/db_xref="taxon:10090"
/chromosome="X"
/map="near Pip"
/tissue_type="liver"
/join(409..509,866..942,1263..1873)
/gene="Bex1"
/product="Bex1 protein"
409..1873
/gene="Bex1"
/note="expressed in brain; X-linked"
1268..1654
/gene="Bex1"
/codon_start=1
/product="Bex1 protein"
/protein_id="AAD47168.1"
/db_xref="GI:5702156"
/translation="MESKDGKVNKNMENDHQBKEEKEKPODITRRPAVALTSEAG
KNCAPRGRRFRVRQPIAHYRWDLMLQRYGEPQGRMRNENVRFGDVRQLMEKLE
RQLSHSLRAVSTDPDHPHDEFCIMP"
BASE COUNT 563 a 554 c 705 g 447 t
ORIGIN

```



```

alignment_scores:
  Quality: 448.50      Length: 129
  Ratio: 3.934         Gaps: 3
  Percent Similarity: 88.372  Percent Identity: 68.217

alignment_block:
  US-09-327-750D-32 x AF097437
  Align seg 1/1 to: AF097437 from: 1 to: 2269

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1268 ATGGAGTCAAAAGAT...CAAGCGTGAAAAATCTCAACATGAGAGATGA 1314
17 AsnGlnGluAsnGluLysGlu.....GlnValAlaAsnLysG 31
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1315 CCATCAGAAAAGAGGAGAGAGGAGAAAGCCACAGATACCATCAGAA 1364

31 lyGluPro...LeuAlaLeuProLeuAspAlaGlyGluTyrCysValPro 46
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1365 GGGAGCCAGCTGTGCCCTCGACCTCGAGGCTGGCAAAACTGTGCACCT 1414

47 ArgGlyAsnArgArgArgPheProValArgGlnProIleLeuGlnTyrAr 63
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1415 AGAGGAGGTCCGAGCGGTTCCGGGTTCGGCAGCCCATCGCTCACTATAG 1464

63 gTrpAspIleMethHisArgLeuGlyGluProGlnAlaArgMetArgGluG 80
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1465 ATGGAGCTGTATGAGAGGTTGGGAGGCCCCAGGAGGATGAGAGAGG 1514

80 luAsnMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeu 96
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1515 AGAAGCTACAGAGGTTTGGGGGTGATGTGAGCAGCTCATGAGAAGCTG 1564

97 ArgGluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 113
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1565 AGGNAAGGAGCTGACCCACAGCTGGGGGGTTAGCACTGACCCGCC 1614

113 oHisHisAspHisAspGluPheCysLeuMetPro 125
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1615 TCATCATGACCACCATGATGAGTTTGCCTCATGCCGCC 1651

seq_name: gb_ro:AF097438

seq_documentation_block:
  LOCUS AF097438 835 bp mRNA ROD 11-APR-1999
  DEFINITION Mus musculus brain expressed x-linked protein 1 (Bex1) mRNA,
  complete cds.
  ACCESSION AF097438
  VERSION AF097438.1 GI:4580589
  KEYWORDS
  SOURCE house mouse.
  ORGANISM Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  TITLE Bex1, a gene with increased expression in parthenogenetic embryos,
  is a member of a novel gene family on the mouse X chromosome
  JOURNAL Hum. Mol. Genet. 8 (4), 611-619 (1999)
  MEDLINE 99172070
  REMARK Erratum: [[published erratum appears in Hum Mol Genet 1999
  May;8(5):943]]
  REFERENCE 2 (bases 1 to 835)
  AUTHORS Brown,A.L. and Kay,G.F.
  TITLE Direct Submission
  JOURNAL Submitted (08-OCT-1998) Cancer Unit, Queensland Institute of
  Medical Research, Herston Rd, Brisbane, Qld 4029, Australia
  FEATURES
    source
      1..835
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      /chromosome="X"

```

TITLE Characterization of genes which exhibit reduced expression during the retinoic acid-induced differentiation of F9 teratocarcinoma cells: Involvement of cyclin D3 in RA-mediated growth arrest

JOURNAL Mol. Cell. Endocrinol. 143 (1-2), 155-166 (1998)

MEDLINE 99021197

REFERENCE 2 (bases 1 to 838)

AUTHORS Faria, T.N., LaRosa, G., Wilen, E., Liao, L. and Gudas, L.J.

TITLE Direct Submission

JOURNAL Submitted (26-FEB-1998) Pharmacology, Cornell University Medical College, 1300 York Avenue, New York, NY 10021, USA

FEATURES

source 1..838

1 /organism="Mus musculus"

2 /db_xref="taxon:10090"

3 /call_line="F9 teratocarcinoma"

4 229..579

5 /note="retinoic acid reduced gene expression"

6 /codon_start=1

7 /product="REX-3"

8 /protein_id="AAC61929.1"

9 /db_xref="GI:3510643"

10 /translation="MENDHKKKEEKPODITRREPAVALISEAGKNCAPRGGRRRF

11 RVQPIAHVNDLMQVGEPOGRMEENVQFGDVRQLMEKLRQLSHSLRAVSTD

12 PPHDHDFECLMP"

13

BASE COUNT 238 a 181 c 251 g 168 t

ORIGIN

alignment_scores:

Quality: 446.50 Length: 129

Ratio: 3.917 Gaps: 3

Percent Similarity: 88.372 Percent Identity: 68.217

alignment_block:

US-09-327-750D-32 x AF051347

Align seg 1/1 to: AF051347 from: 1 to: 838

1 MetGluSerLySGluLyAArgAlaValAsnSerLeuSerMetGluAsnAl 17

193 ATGGATCTCAAGAT...CAAGGCTGAAATCTCAACATGGAGAATGA 239

17 aAsnGlnGluAsnGluGluLySglu.....GlnValAlaAsnLySg 31

240 CCATCAGAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 289

31 lyGluPro...LeuAlaLeuProLeuAspAlaGlyGluTyrCysValPro 46

290 GGGAGGCGCATGTGGCCCTGTCTCCGAGGCTGGCAAAACTGTGCCCT 339

47 ArgGlyAsnArgArgPheProValArgGlnProIleLeuGlnTyrAr 63

340 AGAGGAGGTCGAGCGGTTCCGGTTCGGAGCCCATCGCTACTATAG 389

63 gtrPaspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGlu 80

390 ATGGGACCTGATCAGAGGTTGGGGAGCCCGAGGAGGAGGAGGAGG 439

80 luAsnMetGluArgIleGlyGluValArgGlnLeuMetGluLySleu 96

440 AGAACGTACAGAGTTGGGGTGTGTGATGTGACAGACCTCATGAGAG 489

97 ArgGluLySlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 113

490 AGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 539

113 oHisHisAspHisAspCplupPheCysLeuMetPro 125

540 TCATCATGACACCATCATGATGATGTTTGTCTCATGCC 576

seq_name: gb_ro:BC003254

seq_documentation_block:

LOCUS BC003254 2901 bp mRNA ROD 20-FEB-2001

DEFINITION Mus musculus. Similar to dentatorubral pallidolysian atrophy, clone MGC:5758, mRNA, complete cds.

ACCESSION BC003254

VERSION BC003254.1 GI:13096906

KEYWORDS MGC.

SOURCE house mouse.

ORGANISM Mus musculus.

REFERENCE 1 (bases 1 to 2901)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (20-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Jeffrey Green M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: villalob@bcm.tmc.edu.

Villalob, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

Series: IRAC Plate: 9 Row: 0 Column: 6.

FEATURES

Location/Qualifiers

1..2901

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="MGC:5758 IMAGE:3500522"

/tissue_type="Mammary tumor: C3(1)-Tag model. Infiltrating ductal carcinoma. 5 month old virgin mouse."

/clone_lib="NCI-CGAP_Mam6"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

755..2359

/codon_start=1

/product="Similar to dentatorubral pallidolysian atrophy"

/protein_id="AAH03254.1"

/db_xref="GI:13096907"

/translation="MVPQFLPLTLPGLPKPPHVTPLHPRAPKEHPTTSHOSLOSP PPQFLPLSPPAATGPTTATQIQEPAEYEPPEPPVPAKSPPPKVVDPVS HASQARFNKHLDRGFNSCARSDLYFVPLEGSKLAKRADLVKVRREARQAREEKE REREREKEREREKEREKERSVLAQEGRAPVECPSLGVPVPRPPFEPGSAVATVPP YLGPDTPALRTLSEYARPHVSPGNRNHPYVPLGADPGLLGYNPVALYSSDPAARE RERAREEDLDRLKPGFEVKPSELEPLHGVPGLDPRHGGALOPGPGPLHPFP FHPSLGPLERLALAAPALRDMYSVAERLAERQHAERVAALGNLPLALMLNVT PHHHOHSIHSHLHQDDAIHAASVSHPLIDPLASGSHLRTIPYPAGTLPNPLPH PLHENEVLHQLFAPYRDLPASLAPMSAAHQLOQAHQAOSAELOKRLALEQQOOLHAAH HPLHSVPLPAQEDYYSHLKESDKPL"

BASE COUNT 623 a 958 c 791 g 529 t

ORIGIN

alignment_scores:

Quality: 384.50 Length: 119

Ratio: 3.697 Gaps: 3

Percent Similarity: 87.395 Percent Identity: 65.546

alignment_block:

US-09-327-750D-32 x BC003254

Align seg 1/1 to: BC003254 from: 1 to: 2901

```

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
223 ATGGAGTCCAAAGAT...CAAGGCTGAAATATCTCAACATGGAGATGA 269
17 asnGlnGluAsnGluGluLysGlu.....GlnValAlaAsnLysG 31
270 CCATCAGAAAAGAGGAGAGAGAGAAAGCCCAAGATACCATCAGAA 319
31 lyGluPro...LeuAlaLeuProLeuAspAlaGlyGluTyrcysValPro 46
320 GGGAGCCAGCTGTGGCCCTGACCTCCGAGGCTGCAAAACTGTGCACCT 369
47 ArgGlyAsnArgArgPheProValArgGlnProIleLeuGlnTyrr 63
370 AGAGGAGGTGCGAGGCGGTTCCGGGTTCCGACGCCCATCTCCTACATAG 419
63 gTrrAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGlu 80
420 ATGGGACCTGTATGCAGAGGTTGGGAGCCCGGAGGAGGATGAGAGG 469
80 luAsnMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeu 96
470 AGAAGCTACAGAGGTTTGGGGTGTGTGAGACAGCTCATGGAGAGCTG 519
97 ArgGluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 113
520 AGGAAAGGACGCTGAGCCAGCAGCTCGCGGGGTTAGCACTGACCGCC 569
113 oHisHis 115
570 TCATCAT 576

```

seq_name: gb_sts:G24641

```

seq_documentation_block: 504 bp DNA 31-MAY-1996
LOCUS G24641 human STS WI-11354, sequence tagged site.
DEFINITION G24641
ACCESSION G24641
VERSION G24641.1 GI:1344967
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human STSs derived from sequences in dbEST and the Unigene
collection.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 504)
Whitehead Institute/MIT Center for Genome Research; Physically
Mapped STSs
Unpublished (1995)

```

Contact: Thomas Hudson
 Whitehead Institute/MIT Center for Genome Research
 Whitehead Institute for Biomedical Research
 9 Cambridge Center, Cambridge MA 02142 USA
 Tel: 617 252 1900
 Fax: 617 252 1902
 Email: thudson@genome.wi.mit.edu

Primer A: TTTTTCCTTTTAAAGGTGC
 Primer B: TTACGTGGGTCTCTATTACCA
 STS size: 150
 PCR Profile:

Presoak:
 Denaturation:
 Annealing: 56 degrees C
 Polymerization:
 PCR Cycles: 35
 Thermal Cycler:

Protocol:
 Template: 10 ng

Primer: each 5 pM
 dNTPs: each 4 mM
 Tag Polymerase: 0.025 units/ul
 Total Vol: 20 ul

Buffer:

MgCl2: 1.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 9.3

Derived from dbEST (genbank accession T86927).

Location/Qualifiers

FEATURES
 source 1..504
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 STS 1..150
 primer_bind 1..22
 primer_bind complement(129..150)
 BASE COUNT 134 a 118 c 101 g 144 t 7 others
 ORIGIN

alignment_scores:

Quality: 348.00 Length: 85
 Ratio: 4.703 Gaps: 3
 Percent Similarity: 87.059 Percent Identity: 84.706

alignment_block:

US-09-327-750D-32 x G24641/rev

Align seg 1/1 to reverse of: G24641 from: 1 to: 504

```

44 CysValProArgGly.AsnArgArgArgPhe.ProValArgGlnProIle 59
503 TGTCCCCCTTANAGAAANCCGTAGCGGTTCCCCCTTAGGCANCCCCAC 454
60 LeuGlnTyrrArgTrrAspIleMetHisArgLeuGlyGlu.ProGlnAla 76
453 CGCAGATANAGATGGATATGATGATAGGTTGGAGAACCCACAGGCAA 404
76 rgMetArgGluAsnMetGluArgIleGlyGluGluValArgGlnLeu 92
403 GGATGAGAGAGAGATATGAAAGGATTTGGGAGGAGGTGAGACAGCTG 354
93 MetGluLysLeuArgGluLysGlnLeuSerHisSerLeuArgAlaVal 109
353 ATGGAAAAGCTGAGGAAAAGCAGTTGAGTCANAGTCTCGCGGCGAGTCAG 304
109 rThrAspProProHisHisAspHisAspGluPheCysLeuMetPro 125
303 CACTGACCCCCCTCCACCATGACCATGATGAGTTTTTGTCTTATGCC 255

```


ethanol-precipitation. The cDNAs were ligated to lone-linker LL-Sal3 (include SalI sequence). The cDNAs were purified by phenol/chloroform and separated from free linkers by Centricon 100. Then, cDNAs were amplified by long-range high fidelity PCR using Takara's Ex Taq polymerase. Then, the cDNAs were purified by phenol/chloroform and by Centricon 100. The cDNAs were digested with SalI and NotI enzymes. Then, the cDNAs were size selected by Gibco's Size Fractionation Column. The cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by chemical method. The library was constructed by Xiaohong Wang and Minoru S. H. Ko."

BASE COUNT 145 a 182 c 133 g 182 t
ORIGIN

alignment_scores:
Quality: 646.00 Length: 128
Ratio: 5.168 Gaps: 0
Percent Similarity: 97.656 Percent Identity: 92.969

alignment_block:
US-09-327-750D-33 x AW536404/rev ..

Align seg 1/1 to reverse of: AW536404 from: 1 to: 642

1 MetGluSerLysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspHi 17
|||||
606 ATGAGTCCAAAGATCAAGCGGTGAAATCTCACTGGAGAGTACCA 557
17 sGlnLysGluGluLysGluLysProGlnAspThrIleLysArgG 34
|||||
556 TCACAAAAGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 507
34 LuProValValAlaProThrPheGluAlaGlyLysAsnCysAlaProArg 50
|||||
506 AGCCAGCTGGCGCTCCCTCCGAGGCTGGCAAACTGTGCACCTAGA 457
51 GlyGlyArgArgPheArgValArgGlnProIleSerHisTyArgTr 67
456 GGAGTGCAGGCGGTTCGGGTTCGGAGCCCATCGCTCACTATAGT 407
67 PASLeuMethHisArgValGlyLysProGlnGlyArgMetArgGlu 84
|||||
406 GGACCTGATGACGAGGCTGGGAGGAGGAGGAGGAGGAGGAGGAG 357
84 snValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 100
|||||
356 ACGTACAGAGGTTTGGGGGTGATGTGAGACAGCTCATGGAGAGCTG 307
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 117
|||||
306 GAAAGGACGCTGACCCACAGCCTGGGGCGGTAGCAGTACCCGCTCA 257
117 sHisAspHisArgGluPheCysLeuMetPro 128
256 TCATGACCCACCATGATGAGTTTGCCTCATGCCC 223

seq_name: gb_est2:BG064920

seq_documentation_block: 707 bp mRNA EST 26-JAN-2001
LOCUS BG064920
DEFINITION H3025D01-3 NIA Mouse 15K cDNA clone set Mus musculus cDNA clone
H3025D01 3', mRNA sequence.

ACCESSION BG064920
VERSION BG064920.1 GI:12547483
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 707)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Craniata; Chordata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

34 LuProValValAlaProThrPheGluAlaGlyLysAsnCysAlaProArg 50
|||||
242 ACCAGCTGGCGCTCCGAGGCTGGCAAACTGTGCCTTAGA 291
51 GlyGlyArgArgPheArgValArgGlnProIleSerHisTyArgTr 67
292 GGAGTGCAGGCGGTTCGGGTTCGGAGCCCATCGCTCACTATAGT 341
67 PASLeuMethHisArgValGlyLysProGlnGlyArgMetArgGlu 84
342 GGACCTGATGACGAGGTTGGGAGGAGGAGGAGGAGGAGGAGGAG 391
84 snValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 100
392 ACGTACAGAGGTTTGGGGGTGATGTGAGACAGCTCATGGAGAGCTG 441
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 117
442 GAAAGGACGCTGACCCACAGCCTGGGGCGGTAGCAGTACCCGCTCA 491
117 sHisAspHisArgGluPheCysLeuMetPro 128
492 TCATGACCCACCATGATGAGTTTGCCTCATGCCC 525

seq_name: gb_est1:AW536404

seq_documentation_block: 642 bp mRNA EST 31-AUG-2000
LOCUS AW536404
DEFINITION G0104A11-3 NTA Mouse E7.5 Embryonic Portion cDNA Library Mus
musculus cDNA clone G0104A11 3', mRNA sequence.
ACCESSION AW536404
VERSION AW536404.1 GI:7178821
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 642)
AUTHORS Tanaka, T.S., Jaradat, S.A., Lim, M.K., Kargul, G.J., Wang, X., Grahovac, M.J., Pantano, S., Sano, Y., Piao, Y., Nagaraja, R., Doi, H., Wood, W.H., III, Becker, K.G. and Ko, M.S.H.
TITLE Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
MEDLINE 20381348
COMMENT Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@nslu.gsc.nia.nih.gov
Plate: G0104 row: A column: 11
Seq primer: -21M13 Forward
High quality sequence stop: 642
POLYA=Yes.

FEATURES
Location/Qualifiers
1..642
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G0104A11"
/clone_lib="NIA Mouse E7.5 Embryonic Portion cDNA Library"
/sex="unknown"
/dev_stage="7.5dpc Embryo"
/lab_host="DH10B"
/note="Vector: pSPORT1 (Gibco/BRL Life Technology);
Site1: SalI; Site2: NotI; Total RNAs were extracted from
6 Embryo. The double-stranded cDNA was synthesized by
Gibco's kit with an Oligo(dT) primer (NotI primer-adaptor
from GibcoBRL)
[5'-pGACTAGTCTAGATCGGAGCGGCGCTTTTCTTTT-3']
from 0.5µg of mRNA. The double-stranded cDNAs were
treated with T4 DNA polymerase and purified by

AUTHORS Kargul G.J., Dudekula D.B., Qian Y., Lim, M.K., Jaradat, S.A., Tanaka T.S., Carter, M.G. and Ko, M.S.H.
TITLE Verification and initial annotation of NIA mouse 15K cDNA clone set
JOURNAL Unpublished (2001)
COMMENT Other_ESTs: H3025D01-5
 Contact: George J. Kargul
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@gsun.grc.nia.nih.gov
 This clone set has been freely distributed to the community. Please visit <http://lgsun.grc.nia.nih.gov/cDNA/15k.html> for details.
 Plate: H3025 row: D column: 01
 Seq primer: -21M13 Forward
 High quality sequence stop: 707
 POLYA=Yes

FEATURES Location/Qualifiers
 source
 1..707
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="H3025D01"
 /clone_lib="NIA Mouse 15K cDNA Clone Set"
 /sex="Clones arrayed from a variety of cDNA libraries"
 /dev_stage="Clones arrayed from a variety of cDNA libraries"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; Site1: SalI; Site2: NotI; This clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."
 Mol Genet 7: 1967-1978."

BASE COUNT 154 a 209 c 146 g 198 t
ORIGIN
 alignment_scores
 Quality: 643.00 Length: 128
 Ratio: 5.144 Gaps: 0
 Percent Similarity: 97.656 Percent Identity: 92.188
 alignment_block
 US-09-327-750d-33 x BG064920/rev
 Align seg 1/1 to reverse of: BG064920 from: 1 to: 707
 1 MetGluSerLysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspH1 17
 606 ATGGAGTCCAAAGATCAAGCGTGAAATAATCTCAACATGGAGAATGACCA 557
 17 scInLysLysGluGluLysGluGluLysProGlnAspThrIleLysArgG 34
 556 TCAGAAAAAGGAGGAGGAAGAAACCAAGATACCATCAGAAAGG 507
 34 luProValValAlaProThrPheGluAlaGlyLysAsnCysAlaProArg 50
 506 AGCCAGCTGTGGCCCTGCACCTCCGAGGCTGGCAAAAACCTGCACCTAGA 457
 51 GlyGlyArgArgArgPheArgValArgGlnProIleSerHisTyArgTr 67

|||||
 456 GGAGTCCGACGCGGTTCGGGTTCCGACGCCCATCGCTACATATAGTG 407
 67 pAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluGluA 84
 406 GGACCTGATGACGAGGGTTGGGGAGCCCCCGGGAAGGATGAGAGAGGAGA 357
 84 snValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuArg 100
 356 AGCTACACAGGTTTGGGGTGTGAGACAGCTCATGGAGAGCTGAGG 307
 101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProH1 117
 306 GAAGGACGCTGAGCCACGCGTGGGGCGGTGAGCACTGACCCGCTCA 257
 117 shisAspHisAspGluPheCysLeuMetPro 128
 256 TCATGACCACCATGATGAGTTTGCCTCATGCCC 223
 seq_name: gb_est2:BG277659
 seq_documentation_block:
 LOCUS BG277659 575 bp mRNA EST 21-FEB-2001
 DEFINITION ux47c11.y1 Soares.NMMAX_maxillary_process Mus musculus cDNA clone
 INAMES:3513237 5' similar to TR:Q9RLJ2 Q9RLJ2 BEX1 PROTEIN. ; mRNA
 sequence.
 ACCESSION BG277659
 VERSION BG277659.1 GI:13073183
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE 1 (bases 1 to 575)
 NC-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Other_ESTs: ux47c11.x1
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-r@mail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:1393813
 Seq primer: -40RP from Gibco
 High quality sequence stop: 462.
 Location/Qualifiers
 1..575
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:3513237"
 /clone_lib="Soares.NMMAX_maxillary_process"
 /tissue_type="maxillary process"
 /lab_host="DH10B (phage-resistant)"
 /note="Vector: pT73D-pac (Pharmacia) with a modified
 polylinker; Site1: NotI; Site2: EcoRI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTACCAATCTGAAGTGGAGCGCGCGGCTTTTGTATTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 156 a 136 c 201 g 81 t 1 others
 ORIGIN
 alignment_scores
 Quality: 633.00 Length: 128
 Ratio: 5.105 Gaps: 0
 Percent Similarity: 96.875 Percent Identity: 91.406
 alignment_block:

US-09-327-750D-33 x BG277659

Align seg 1/1 to: BG277659 from: 1 to: 575

```

1 MetGluSerLysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspH1 17
|||||
192 ATGGAGTCCAAAGATCAAGCGGTGAAAAATCTCAACATGGAGAATGACCA 241
|||||
17 sGlnLysLysGluGluLysGluLysProGlnAspThrIleLysArgG 34
|||||
242 TCAGAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 291
|||||
34 luProValValAlaProThrPheGluAlaGlyLysAsnCysAlaProArg 50
|||||
292 AGCCAGCTGTGGCCCTGACCTCCGAGGCTGGCAAAACTGTGCACCTAGA 341
|||||
51 GlyGlyArgArgPheArgValArgGlnProIleSerHisTyArgTr 67
|||||
342 GGAGTGCAGGCGGTTCGGGGTTCGGCAGCCCATCGCTCACTATAGATG 391
|||||
67 pAspLeuMethHisArgValGlyGluProGlnGlyArgMetArgGluAla 84
|||||
392 GCACCTGATCCAGAGGTTGGGAGGCCGCCAGGGAAGGATGAGAGGAGA 441
|||||
84 snValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuArg 100
|||||
442 ACGTACAGAGTTGGGGTGATGTGAGACAGCTCATGGAGAGCTGAGG 491
|||||
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProH1 117
|||||
492 GAAAGGAGCTGTAGCCACAGCTCGGGGGTTAGCACGTGACCCGCCCTCA 541
|||||
117 sHisAspHisArgGluPheCysLeuMetPro 128
|||||
542 TCATGACCACTGATGAGTTGTGCTCATGCCCC 575
|||||

```

seq_name: gb_est2:BG870503

```

seq_documentation_block: 756 bp mRNA EST 29-MAY-2001
LOCUS BG870503 602791422F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4922644 5',
DEFINITION mRNA sequence.
ACCESSION BG870503
VERSION BG870503.1 GI:14221043
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 756)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10842 row: 1 column: 05
High quality sequence stop: 756.
Location/Qualifiers
1..756
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4922644"
/clone_lib="NCI_CGAP_SG2"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:"

```

FEATURES

```

source
1..756
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4922644"
/clone_lib="NCI_CGAP_SG2"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:"

```

NOTI; Site_2; SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 207 a 164 c 230 g 155 t

ORIGIN

alignment_scores:
Quality: 633.00 Length: 129
Ratio: 5.064 Gaps: 1
Percent Similarity: 96.899 Percent Identity: 92.248

alignment_block:

US-09-327-750D-33 x BG870503

Align seg 1/1 to: BG870503 from: 1 to: 756

```

1 MetGluSerLysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspH1 17
|||||
165 ATGGAGTCCAAAGATCAAGCGGTGAAAAATCTCAACATGGAGAATGACCA 214
|||||
17 sGlnLysLysGluGluLysGluLysProGlnAspThrIleLysArgG 34
|||||
215 TCAGAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 264
|||||
34 luProValValAlaProThrPheGluAlaGlyLysAsnCysAlaProArg 50
|||||
265 AGCCAGCTGTGGCCCTGACCTCCGAGGCTGGCAAAACTGTGCACCTAGA 314
|||||
51 GlyGlyArgArgPheArgValArgGlnProIleSerHisTyArgTr 67
|||||
315 GGAGTGCAGGCGGTTCGGGGTTCGGCAGCCCATCGCTCACTATAGATG 364
|||||
67 pAspLeuMethHisArgValGlyGluProGlnGlyArgMetArgGluGlu 83
|||||
365 GCACCTGATGACAGAGGTTGGGAGCCCCAGGGAAGGATGACGAGGAG 414
|||||
84 AsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuArg 100
|||||
415 AACGTACAGAGGTTGGGGGTGATGTGAGACAGCTCATGGAGAAGCTGAG 464
|||||
100 gGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProH1 117
|||||
465 GGAAGGAGGAGTGCACACAGCTTCGGGGGTGAGCAGCTGACCTGACCCGCC 514
|||||
117 sHisAspHisArgGluPheCysLeuMetPro 128
|||||
515 ATCATGACCACTGATGAGTTGTGCTCATGCCCC 549
|||||

```

seq_name: gb_est1:BB654459

```

seq_documentation_block: 499 bp mRNA EST 06-SEP-2000
LOCUS BB654459 602791422F1 NCI_CGAP_SG2 Mus musculus cDNA clone
DEFINITION UI-M-AJI-aha-f-10-0-UI.r1 NIH_BMAP_MOB_N Mus musculus cDNA clone
UI-M-AJI-aha-f-10-0-UI 5', mRNA sequence.
ACCESSION BB654459
VERSION BB654459
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 499)
Normalizaton and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704477
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301-443-1706

```

Fax: 301 443 9890

Email: mEst@mail.nih.gov

cDNA Library Preparation: M. B. Soares Lab Clone Distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements

Seq primer: M13 Reverse.

Location/Qualifiers

FEATURES

source

1. 499
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-AJ1-aha-f-10-0-UI"
 /clone_lib="NIH_BMAP_MOB_N"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT730-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_MOB_N library is a normalized library constructed from mouse olfactory bulbs. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dr track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories."

146 a 108 c 179 g 66 t

BASE COUNT
ORIGIN

alignment_scores:

Quality: 615.00 Length: 123
 Ratio: 5.125 Gaps: 0
 Percent Similarity: 97.561 Percent Identity: 92.683

alignment_block:

US-09-327-750D-33 x BE654459

Align seg 1/1 to: BE654459 from: 1 to: 499

1 MetGluSerLysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspH1 17
 |||||
 131 ATGGAGTCCAAAGATCAAGCGGTGAAAAATCTCAATGAGGAGTACCA 180
 17 sGlnLysLysGluGluLysGluLysProGlnAspThrIleLysArgG 34
 |||||
 181 TCAGAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 230
 34 luProValValAlaProThrPheGluAlaGlyLysAsnCysAlaProArg 50
 |||||
 231 ACCGAGCTGTGCCCTGACCTCCAGAGCTGGCAAAATCTGGACCTAG 280
 51 GlyGlyArgArgArgPheArgValArgGlnProIleSerHisTyrArgTr 67
 |||||
 281 GGAGTCCAGCGGTTCCGGGTTCCGAGCCATCGCTCACTATAGATG 330
 67 pasLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluGlu 84
 |||||
 331 GGACCTGATGACAGGGTTGGGGAGCCCGAGGAGGAGGAGGAGGAG 380
 84 snValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuArg 100
 |||||
 381 ACGTACAGAGGTTTGGGGGTGATGTGAGACAGCTCATGGAGAGCTGAG 430
 101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH1 117
 |||||
 431 GAAAGGACAGCTGAGCCAGAGCTCGGGCGGTTAGCAGTACCGCGCTCA 480
 117 shisAspHisAspGlu 123
 |||||
 481 TCATCACCACCATGATGAG 499

seq_name: gb_est1:AW536974

seq_documentation_block:

LOCUS AW536974 577 bp mRNA EST 31-AUG-2000
 DEFINITION GO110H10-3 NIA Mouse E7.5 Embryonic Portion cDNA Library Mus
 musculus cDNA clone GO110H10 3', mRNA sequence.

ACCESSION AW536974.1 GI:7179391

VERSION EST.

KEYWORDS house mouse.

SOURCE Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 577)

REFERENCE

AUTHORS

Tanaka, J.T., Jaradat, S.A., Lim, M.K., Kargul, G.J., Wang, X., Grahovac,
 M.J., Pantano, S., Sano, Y., Piao, Y., Nagara, R., Doi, H., Wood, W.H.
 III, Becker, K.G. and Ko, M.S.H.

Genome-wide expression profiling of mid-gestation placenta and

embryo using a 15,000 mouse developmental cDNA microarray

Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)

20381348

MEDLINE

COMMENT

Contact: George J. Kargul
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@igsn.grc.nia.nih.gov

Plate: GO110 row: H column: 10

Seq primer: -21M13 Forward

High quality sequence stop: 577

POLYA=Yes.

FEATURES

source

Location/Qualifiers

1. 577
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="GO110H10"
 /sex="unknown"
 /dev_stage="7.5dpc Embryo"
 /lab_host="DH10B"
 /note="Vector: pSPORT1 (Gibco/BRL Life Technology);
 Site_1: SalI; Site_2: NotI; Total RNAs were extracted from
 6 Embryo. The double-stranded cDNA was synthesized by
 Gibco's kit with an oligo(dT) primer [NotI primer-adaptor
 from GibcoBRL]
 [5'-pGACTAGTCTTAGATCGGAGCGCCCTTTT-3']
 from 0.5ug of mRNA. The double-stranded cDNAs were
 treated with T4 DNA polymerase and purified by
 ethanol-precipitation. The cDNAs were ligated to
 Lone-linker LI-Sal3 (include SalI sequence). The cDNAs
 were purified by phenol/chloroform and separated from
 free linkers by Centricon 100. Then, cDNAs were amplified
 by long-range high fidelity PCR using Takara's Ex Taq
 polymerase. Then, the cDNAs were purified by
 phenol/chloroform and by Centricon 100. The cDNAs were
 digested with SalI and NotI enzymes. Then, the cDNAs were
 size selected by Gibco's Size Fractionation Column. The
 cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
 vector. The DH10B E. coli host was transformed with the
 ligation mixture by chemical method. The library was
 constructed by Xiaohong Wang and Minoru S. H. No."

137 a 160 c 126 g 154 t

BASE COUNT
ORIGIN

alignment_scores:

Quality: 599.00 Length: 118
 Ratio: 5.209 Gaps: 0
 Percent Similarity: 97.458 Percent Identity: 93.220

alignment_block:

US-09-327-750D-33 x AW536974/rev

Align seg 1/1 to reverse of: AW536974 from: 1 to: 577

```

11 LeuAsnMetGluAsnAspHisGlnLysLysGluLysGluLysGluLysPr 27
576 CTCACATGGAGTAATGACCATCAGAAAAGGAGGAGGAAGAAAGGCC 527
27 oGlnAspThrIleLysArgGluProValAlaProThrPheGluAla 44
526 ACAAGATACCATCAGAAAGGAGCCAGCTGTGGCCCTGACCTCCGAGGCTG 477
44 LysAsnCysAlaProArgGlyGlyArgArgPheArgValArgGln 60
476 GCAAAACCTGTCCACTAGAGAGGTCCAGCGGTTCGGGTTCGCGAG 427
61 ProIleSerHisTyrArgTrpAspLeuMetHisArgValGlyGluProG1 77
426 CCCATCGCTCACTATAGATGGACCTGATGACAGAGGTGGGAGGCCCA 377
77 nGlyArgMetArgGluGluAsnValGlnArgPheGlyGluAspMetArg 94
376 GGGAGGATGAGAGAGGAGACGATACAGAGTTTGGGGGTGTGTGAGAC 327
94 InLeuMetGluLysLeuArgGluArgGlnLeuSerHisSerLeuArgAla 110
326 AGCTCATGAGAAGCTGAGGGAAGGACGCTGAGCCACAGCTGCGGGCG 277
111 ValSerThrAspProHisHisAspHisHisAspGluPheCysLeuMe 127
276 GTTAGCAGTGCACCCGCTCATCATGACACCATGATGAGTTTGCCTCAT 227
127 tPro 128
226 GCCC 223

```

seq_name: gb_est2.BF607762

```

seq_documentation_block: 785 bp mRNA 01-APR-2001
LOCUS BF607762
DEFINITION MY1_000704 Mouse 9-day fetus cDNA library ICRP522 Mus musculus
cDNA clone ICRP522A1947 5', mRNA sequence.
ACCESSION BF607762
VERSION BF607762.1 GI:13504254
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 785)
Yahyawi, M., Hennig, S., Neidhardt, L., Radelof, U., Hermann, B. G.,
Lehrach, H. and O'Brien, J.
Detection of a high number of novel genes in a 9-day mouse embryo
cDNA library normalised by oligonucleotide fingerprinting
JOURNAL Unpublished (2001)
COMMENT Contact: Hennig S
Laboraty 123, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr. 63-73, D-14195 Berlin, Germany
Tel.: +49 30 8413 1612
Fax: +49 30 8413 1380
Email: hennig@molgen.mpg.de
EST's are made from clones being representatives of clone clusters.
clone clusters were calculated from oligonucleotide fingerprints.
PCR Primers
FORWARD: 5'-GAGCTATTCCAGAGTAGTGA-3'
BACKWARD: 5'-TAATGAGCTCACTATAGGG-3'
Seq primer: 5'-ATTAGTGACACTAG-3'
High quality sequence stop: 785.
Location/Qualifiers
1..785
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="ICRP522A1947"
/clone_lib="Mouse 9-day fetus cDNA library ICRP522"

```

FEATURES
source

```

/tissue_type="whole embryo"
/dev_stage="embryonic 9-day"
/lab_host="E.coli, XL1 blue"
/notes="vector: PSVSPORT1; Site_1: NotI; Site_2: SalI;
Library preparation by oligo-dT priming of RNA. Clones can
be ordered from the Resource Center in Berlin,
http://www.rzpd.de."

```

BASE COUNT 190 a 188 c 233 g 153 t 21 others
ORIGIN

alignment_scores:
Quality: 554.00 Length: 128
Ratio: 4.617 Gaps: 4
Percent similarity: 93.750 Percent identity: 87.500

alignment_block:
US-09-327-750D-33 x BF607762

Align seg 1/1 to: BF607762 from: 1 to: 785

```

1 MetGluSerLysAspGlnGlyAlaLysAsnMetGluAsnAspH1 17
259 ATGGAGTCCAAAGATCAAGCGCTGAAAAATCTCAACATGGAGATGACCA 308
17 sGlnLysLysGluGluLysGluLysProGlnAspThrIleLysArg 34
309 TCAGAAAAGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 358
34 LuProValValAlaProThrPheGluAlaGlyLysAsnCysAlaProArg 50
359 AGCCAGCTGTGGCCCTGACCTCCGAGGCTGGCAAAACTGTGCACCTAGA 408
51 GlyGlyValArgArgPheArgValArgGlnProIleSerHisTyrArgT 67
409 GGAGGTGCGAGCGGTTCGGGTTCGGCAGCCATCGCTACTATAGATG 458
67 pAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluGlu 84
459 GGACCTGTGTCAGAGGGTTCGGGAGAGAGAGAGAGAGAGAGAGAGAG 508
84 snValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuAr 100
509 ACGTACAGAGTTTGGCGGTGATGTGAGACAGCTCATGGGGAGCTGAG 558
100 gGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAsp.ProPro 116
559 GGAAGGACACCTGAACACAGAGCTGGGGCGGTAGCAGCTGACCCGCGCT 608
117 HisHisAspHisHisAspGluPhe 124
609 CATTATGACCAACCATGGATGAGTTT 634

```

seq_name: gb_est2.BG228077

```

seq_documentation_block: 587 bp mRNA 08-FEB-2001
LOCUS BG228077
DEFINITION ux47c11.x1 Soares_NMMAX_maxillary_process Mus musculus cDNA clone
IMAGE:3513237 3' similar to TR:Q9RIJ2 Q9RIJ2 BEX1 PROTEIN. ; mRNA
sequence.
ACCESSION BG228077
VERSION BG228077.1 GI:12715592
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 587)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap
Tumor Gene Index
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.

```


456 ACGTACAGAGGTCTGGGGTGATGTGAGACAGCTCATGGAGAACTCAGG 505

101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHI 117

506 GAAAGGAGCTTATGACACAGCTCGGGGGGTGTAGCACTGACNCGCTCA 555

117 sHisAspHisAspGluPheCysLeuMetPro 128:

556 TCATGACTACCATGATTAGTTTTCCTCATGCC 589

seq_name: gb_est1.BE291071

seq_documentation_block: 583 bp mRNA EST 13-JUL-2000
LOCUS BE291071
DEFINITION 60108631F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3500522 5',
mRNA sequence.

ACCESSION BE291071

VERSION BE291071.1 GI:9172545

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

REFERENCE 1 (bases 1 to 583)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>;
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

Tissue Procurement: Jeffrey Green M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Clone Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L1A8559 row: e column: 03

High quality sequence start: 5

High quality sequence stop: 503.

Location/Qualifiers

1..583

FEATURES

source

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone_image="3500522"

/clone_lib="NCI_CGAP_Mam6"

/sex="female, virgin"

/tissue_type="infiltrating ductal carcinoma"

/dev_stage="5 months"

/lab_host="DH10B"

/note="organ: mammary; Vector: pCMV-SPORT6; Site: 1: SalI;

Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Library constructed by Life Technologies. Investigator

providing samples: Jeffrey Green, M.D., NIH

165 a 136 c 201 g 81 t

BASE COUNT

ORIGIN

alignment_scores:

Quality: 505.50 Length: 120

Ratio: 4.554 Gaps: 2

Percent Similarity: 92.500 Percent Identity: 83.333

alignment_block:

US-09-327-750d-33 x BE291071

Align seg 1/1 to: BE291071 from: 1 to: 583

1 MetGluSerLysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspHI 17

208 ATGGAGTCCAAAGATCAAGCGGTGAAATCTCAACATGGAGAAATGACCA 257

17 sGlnLysLysGluLysGluLysProGlnAspThrIleLysArgG 34

258 TCAGAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 307

34 LuProValValAlaProThrPheGluAlaGlyLysAsnCysAlaProArg 50

308 ACCAGCTGTGGCCCTGACCTCCGAGGCTGGCAAAATCTGCACCTAGA 357

51 GlyGlyArgArgPheArgValArgGlnProIleSerHisTyArgTr 67

358 GGAGGTGCAGAGGGGTTCCGGGTTCGGCAGCCCATCGCTCACTATAGTG 407

67 pAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluGua 84

408 GGACTCATCGAGAGGTTGGGGAGCCCATCGAAGGATGAGAGAGGAGA 457

84 snValGlnAArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuArg 100

458 ACGTACAGAGGTTAGGGGGTGTGTGAGACAGCTCATGGAGAAGCTGAGG 507

101 GluArgGlnLeuSerHis...SerLeuArgAlaVal.SerThrAspPro 116

508 GAAAGGCTGTGAGCCACAGACATCCGAGGCGGTTAGACAATGACCCGA 557

116 rOHsHis 118

558 CTCATCAT 565

seq_name: gb_gss:A2936393

seq_documentation_block: 637 bp DNA GSS 26-APR-2001

LOCUS A2936393

DEFINITION 2M0193105F Mouse 10kb plasmid UUGC2M library Mus musculus genomic

clone UUGC2M0193105 F, DNA sequence.

ACCESSION A2936393

VERSION A2936393.1 GI:13794974

KEYWORDS GSS

SOURCE house mouse.

ORGANISM

REFERENCE 1 (bases 1 to 637)

AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0193 row: L column: 05

Seq primer: CGTTGTAAACGACGCGCAGT

Class: plasmid ends

High quality sequence stop: 637.

Location/Qualifiers

1..637

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC2M0193105"

/clone_lib="Mouse 10kb plasmid UUGC2M library"

/sex="Female"

/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (female) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g14732114|9b|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 131 a 137 c 132 g 176 t 1 others
ORIGIN

alignment_scores:
Quality: 493.50 Length: 117
Ratio: 4.569 Gaps: 1
Percent Similarity: 92.308 Percent Identity: 84.615

alignment_block:

US-09-327-750D-33 x A2936393/rev ..

Align seg 1/1 to reverse of: A2936393 from: 1 to: 637

1 MetGluSerLys... AspGlnGlyAlaLysAsnLeuAsnMetGluAsnAs 16
|||||
382 ATGGAGTCCCAAGTGGAAACAGGCGTGAATAATCTCAACATGGAGAATGA 333
16 pHisGlnLysLysGluGluGluLysProGlnAspThrIleLysA 33
|||||
332 CCATCAGGAAAGGAGGAAAGGAAAGGAAAGCCAGCATCTAGCAAAA 283
33 rGluProValValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49
|||||
282 GGGATCCGATTGGGCGCTTCGAGAGCTGAGACTACTACGTGCCT 233
50 ArgGlyClyArgArgPheArgValArgGlnProIleSerHisTyrAr 66
|||||
232 AGAGGAGTCCAGGCGGTTCGGGTTCGGCAGCCATCGTGCCTACAG 183
66 gTTPAspLeuMetHisArgValGluProGlnGlyArgMetArgGluG 83
|||||
182 ATGGAGCTGATCATAGAGTGGGAGTGGGAGCCCGAGGAGGATGAGAGG 133
83 luAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99
|||||
132 AGAACGTACAGAGTTTGGGGATGATGTGAGACAGCTCATGGAGAAGCTN 83
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
|||||
82 GG.GAAAGGCGCTGAGCCACACACCTTCGGCGCGGTTCACACTGACCCGCC 34

116 o 116

33 T 33

seq_name: gb_est2:BG669326

seq_documentation_block:

LOCUS BG669326 412 bp mRNA EST 30-APR-2001
DEFINITION DRNAEB11 Rat DRG library Rattus norvegicus cDNA clone DRNAEB11
5', mRNA sequence.
ACCESSION BG669326
VERSION BG669326.1 GI:13891248
KEYWORDS EST
SOURCE Norway rat.
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus
1 (bases 1 to 412)
Xiao, H.S., Han, Z.G., Zhang, F.X., Huang, Q.H., Lu, Y.J., Bao, L., Fu, G.,
Guo, C., Yan, Q., Jin, S.X., Zhu, Z.D., Xu, X.R., Li, W.G., Chen, Z. and
Zhang, X.

Distinct gene expression profiles of rat dorsal root ganglion
induced by peripheral nerve axotomy

Unpublished (2001)

JOURNAL

COMMENT

Contact: Zhang Xu

Laboratory of Sensory System

Institute of Neuroscience

320 Yue Yang Road, Shanghai 200031, P.R.China

Tel: 86-21-64748700-121

Fax: 86-21-64711346

Email: xu.zhang@ion.ac.cn

This clone is also available at Chinese National Human Genome
Center at Shanghai, 351 Guo Shouling Road, Zhangjiang Hi-Tech Park,
Pudong New Area, P.R.China. Please contact with Zhang Xu
(xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)

PCR Primers

BACKWARD: T7

Seq primer: T3

POLYA-No.

Location/Qualifiers

1..412

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="DRNAEB11"

/clone_lib="Rat DRG Library"

/sex="male"

/tissue_type="dorsal root ganglion"

/dev_stage="adult"

BASE COUNT 124 a 83 c 130 g 75 t

ORIGIN

alignment_scores:

Quality: 489.50 Length: 126
Ratio: 4.450 Gaps: 2
Percent Similarity: 87.302 Percent Identity: 76.984

alignment_block:

US-09-327-750D-33 x BG669326 ..

Align seg 1/1 to: BG669326 from: 1 to: 412

1 MetGluSerLys... AspGlnGlyAlaLysAsnLeuAsnMetGluAsnAs 16
|||||
37 ATGGAGTCCCAAGTGGAAACAGGCGTGAATAATCTCAACATGGAGAATGA 86
16 pHisGlnLysLysGluGluGluLysProGlnAspThrIleLysA 33
|||||
87 CCATCAGGAAAGGAGGAAAGGAAAGGAAAGCCAGCATCTAGCAAAA 136
33 rGluProValValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49
|||||
137 GGGAGCGGTTGTAGCCCTTCCTTTTGAAGCTGGAGAAATACCTACGTGCCT 186
50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyrAr 66
|||||
187 ATAGGAAGTCCAGAGCGTTCCCGGTTTGGCAGCCCATTTGTCACATATAT 236
66 g.TTPAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGlu 82
237 GATGGACCTCATCGCAAGAGTTGGAGAGCCCGCAGGAGGATGAGAGAA 286
83 GluAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLe 99
|||||
287 GAGATGTGCAAGAGTTTGGGAGGATATGAGACAGCTCATGTGAATTT 336


```

/db_xref="taxon:10090"
/clone="G0106H08"
/clone_lib="NIA Mouse E7.5 Embryonic Portion cDNA Library"
/dev_stage="7.5dpc Embryo"
/lab_host="DH108"
/sex="unknown"
/notes="Vector: pSPORT1 (Gibco/BRL Life Technology);
Site 1: SalI; Site 2: NotI; Total RNAs were extracted from
6 Embryo. The double-stranded cDNA was synthesized by
Gibco's kit with an oligo(dT) primer [NotI primer-adaptor
from GibcoBRL]
[5'-pGACTAGTCTAGATCGCGCCGCGCCCTTTT-3']
from 0.5µg of mRNA. The double-stranded cDNAs were
treated with T4 DNA polymerase and purified by
ethanol-precipitation. The cDNAs were ligated to
Lone-linker LL-Sal3 (include SalI sequence). The cDNAs
were purified by phenol/chloroform and separated from
free linkers by Centricon 100. Then, cDNAs were amplified
by long-range high fidelity PCR using Takara's Ex Taq
polymerase. Then, the cDNAs were purified by
phenol/chloroform and by Centricon 100. The cDNAs were
digested with SalI and NotI enzymes. Then, the cDNAs were
size selected by Gibco's Size Fractionation Column. The
cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
vector. The DH108 E. coli host was transformed with the
ligation mixture by chemical method. The library was
constructed by Xiaohong Wang and Minoru S. H. Ko."
BASE COUNT      151 a      151 c      117 g      130 t
ORIGIN

```

```

alignment_scores:
  Quality: 472.00      Length: 97
  Ratio: 5.187        Gaps: 0
  Percent Similarity: 93.814      Percent Identity: 88.660

alignment_block:
US-09-327-750D-33 x AW536634/rev

Align seg 1/1 to reverse of: AW536634 from: 1 to: 549

32  LysArgGluProValValAlaProThrPheGluAlaGlyLysAsnCysAl 48
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
541 AAAGGGATCCGATTGTGGCCCTTCGAGCTGGAGACTACTACGT 492

48  aProArgGlyArgArgPheArgValArgGlnProIleSerHist 65
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491 GCCTAGAGAGGTCGACGCGGTTCCGGGTTCCGACGCCATCGTGCAC 442

65  yArgTrpAspLeuMethHisArgValGlyGluProGlnGlyArgMetArg 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
441 ACAGATGGGACCTGATGCTAGGTTGGGAGGCCGCCAGGGAAGGATGAGA 392

82  GluGluAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLy 98
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
391 GAGGAGACCTACAGAGGTTTGGGGATGATGTGAGACAGCTCATGGAGAA 342

98  sLeuArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAsp 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
341 GCTGAGGGAAGGACGAGTGGCCACAGCTCGGGCGGTTAGCACCTGACC 292

115 roProHisHisAspHisHisAspGluPheCysLeuMetPro 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
291 CGCCTCATCATGACCACCATGATGAGTTTGCCTTATGCC 251

```


HYPOTHETICAL: NO
ANTI-SENSE: NO

LICANT: Bowman, Cindy G.
 LICANT: Dawson, John L.
 LICANT: Dunder, Erik M.

APPLICANT: Dundee, Erik H.
; APPLICANT: Pace, Gary M.

APPLICANT: Suttle, Janet L.
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/951,715A
FILING DATE: 25-SEP-1992
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/772,027
FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: S-18805/A/CGC 1577/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8615
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3624 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3621
OTHER INFORMATION: /product= "Full-length, maize
optimized cryIB"
OTHER INFORMATION: /note= "Disclosed in Figure 6."
US-07-951-715A-6

```

alignment_scores:
  Quality: 93.00      Length: 116
  Ratio: 1.550      Gaps: 4
  Percent Similarity: 51.724      Percent Identity: 26.724

alignment_block:
US-09-327-750D-33 x US-07-951-715A-6  ..

Align seg 1/1 to: US-07-951-715A-6 from: 1 to: 3624

14  GluasnAspHisClnIyLysGluGluLysGluGluLysProGlnAspGth 30
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1026  GAACACGAGCAGCCACCGGGGCCCAACACACGACGATCAACCCGTCACCC 1075
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
30  rIleLysArgGluProValAlaProThrPheGluAlaGlyLysAsnC 47
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1076  TGCCTTCGCGCAGCCGCGAGCTGTACCGCAGGAGAGTACCGCGCGGTG 1125
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
47  ysAla.....ProArgGlyGlyArgArg 55
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1126  CTGCTGGGGCATCTACCTGGAGCCCATCCAGCGCGTGCCACCGTCG 1175
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
56  PheargValArgGlnProIleSerHisYrArgTrpAspLeuMetHisAr 72
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1176  CTTCAACTTCACCAACCCGACCAAGATCATGAGGA.....CGCGCGCACC 1219
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

72 gValgylgIuproGInglYArMetArgGluasnValGlnArgpHeG 89
1220 CCAACTACACCA.....GCCTACGAGAGCCCGCGCTCGACCTGAAG 1263
89 lyGluAspMetArgGInLeuMetGluLysLeuArgGluArgGlnLeuSer 105
1264 GACAGCGAGACCGAGCTGCCCGCCGACGACACCGAGCGCCCAA..... 1307
106 HisSerLeuAlaValSerThrAspProProHisHisAspHisHis 121
1308 .....CTACGAGAGTACGACCGCGCTGAGCCACATCGGCATCAT 1349
seq_name: /cgn2_6/ptodate/2/1na/5B_COMB.seq:US-08-459-448A-6
seq_documentation_block:
; Sequence 6, Application US/08459448A
; Patent No. 5859336
; GENERAL INFORMATION:
; APPLICANT: Koziel, Michael G.
; APPLICANT: Desai, Nalini M.
; APPLICANT: Lewis, Kelly S.
; APPLICANT: Kramer, Vance C.
; APPLICANT: Warren, Gregory W.
; APPLICANT: Evola, Stephen V.
; APPLICANT: Crossland, Lyle D.
; APPLICANT: Wright, Martha S.
; APPLICANT: Merlin, Ellis J.
; APPLICANT: Launis, Karen L.
; APPLICANT: Rothstein, Steven J.
; APPLICANT: Bowman, Cindy G.
; APPLICANT: Dawson, John L.
; APPLICANT: Dunder, Erik M.
; APPLICANT: Pace, Gary M.
; APPLICANT: Suttie, Janet L.
; TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
; TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESSES:
; ADDRESSES: No. 5859336artis Corporation
; STREET: Patent & Trademark Dept., 520 White Plains
; STREET: Rd., POB 2005
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-9005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,448A
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/951,715
; FILING DATE: 25-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/772,027
; FILING DATE: 04-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: 40403
; REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8582
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3624 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3621
; OTHER INFORMATION: /product= "Full-length, maize
; OTHER INFORMATION: Optimized cryiB"
; OTHER INFORMATION: /note= "Disclosed in Figure 6."
;
; US-08-459-448A-6
;
; alignment_scores:
; Quality: 93.00 Length: 116
; Ratio: 1.550 Gaps: 4
; Percent Similarity: 51.724 Percent Identity: 26.724
;
; alignment_block:
; US-09-327-750D-33 x US-08-459-448A-6
;
; Align seg 1/1 to: US-08-459-448A-6 from: 1 to: 3624
;
; 14 GluAsnAspHisGlnLysGluLysGluLysGluLysProGlnAspTh 30
; 1026 GAACACGAGCAGCCAGCGCGCCACCAACACGAGTCAACCCCGTGACCC 1075
;
; 30 rIleLysArgGluProValValAlaProThrPheGluAlaGlyLysAsnC 47
; 1076 TGCCTTTCGCGCAGCGCGAGCTGTACCGCAGCGAGTACCGCGCGTG 1125
;
; 47 ysAla.....ProArgGlyGlyArgArg 55
; 1126 CTGCTGTGGGCGATCTACCTGGAGCCCATCCACGCGTGCACCGTGGC 1175
;
; 56 PheArgValArgGlnProIleSerHisTyrArgTrpAspLeuMetHisar 72
; 1176 CTTCACTTCACCAACCCCGCAGCAATCAGCA.....CGCGGCGACCG 1219
;
; 72 gValcGlyGluProGlnGlyArgMetArgGluGluAsnValGlnArgPheG 89
; 1220 CCAACTACAGCA.....GCCCTAGAGAGCGCGCGCTGCAGCTGAAG 1263
;
; 89 lyGluAspMetArgGlnLeuMetGluLysLeuArgGluArgGlnLeuSer 105
; 1264 GACAGCAGACCGAGCTGCCCCCGAGACCCAGCCGAGCGGCCCA..... 1307
;
; 106 HisSerLeuArgAlaValSerThrAspProProHisHisAspHisHis 121
; 1308 .....CTACGAGAGCTACAGCCAGCGCGCTGAGCCACATCGGCATCAT 1349
;
; seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-459-595A-6
;
; seq_documentation_block:
; ; Sequence 6, Application US/08459595A
; ; Patent No. 6018104
; ; GENERAL INFORMATION:
; ; APPLICANT: Koziel, Michael G.
; ; APPLICANT: Desai, Nalini M.
; ; APPLICANT: Lewis, Kelly S.
; ; APPLICANT: Kramer, Vance C.
; ; APPLICANT: Warren, Gregory W.
; ; APPLICANT: Evola, Stephen V.
; ; APPLICANT: Crossland, Lyle D.
; ; APPLICANT: Wright, Martha S.
; ; APPLICANT: Merlin, Ellis J.
; ; APPLICANT: Launis, Karen L.
; ; APPLICANT: Rothstein, Steven J.
; ; APPLICANT: Bowman, Cindy G.
; ; APPLICANT: Dawson, John L.
; ; APPLICANT: Dunder, Erik M.
; ; APPLICANT: Pace, Gary M.
; ; APPLICANT: Suttie, Janet L.

```

```

;
; TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
; TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6018104artis Corporation
; STREET: Patent & Trademark Dept., 520 White Plains
; STREET: Rd., POB 2005
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-9005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,595A
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/951,715
; FILING DATE: 25-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/772,027
; FILING DATE: 04-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: 40403
; REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8582
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3624 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3621
; OTHER INFORMATION: /product= "Full-length, maize
; OTHER INFORMATION: Optimized cryiB"
; OTHER INFORMATION: /note= "Disclosed in Figure 6."
;
; US-08-459-595A-6
;
; alignment_scores:
; Quality: 93.00 Length: 116
; Ratio: 1.550 Gaps: 4
; Percent Similarity: 51.724 Percent Identity: 26.724
;
; alignment_block:
; US-09-327-750D-33 x US-08-459-595A-6
;
; Align seg 1/1 to: US-08-459-595A-6 from: 1 to: 3624
;
; 14 GluAsnAspHisGlnLysGluLysGluLysGluLysProGlnAspTh 30
; 1026 GAACACGAGCAGCCAGCGCGCCACCAACACGAGTCAACCCCGTGACCC 1075
;
; 30 rIleLysArgGluProValValAlaProThrPheGluAlaGlyLysAsnC 47
; 1076 TGCCTTTCGCGCAGCGCGAGCTGTACCGCAGCGAGTACCGCGCGTG 1125
;
; 47 ysAla.....ProArgGlyGlyArgArg 55
; 1126 CTGCTGTGGGCGATCTACCTGGAGCCCATCCACGCGTGCACCGTGGC 1175

```

```
56 PheargValArgGlnProIleSerHisTyArgTrpAspLeuMetHisar 72
      ::::::::::::::::::::|||
1176 CTTCAACTTCAACACCCAGACATACCGA.....CCGGGACCG 1219
      ::::::::::::::::::::|||
72 gValGlyGluProGlnGlyArgMetArgGluGluAsnValGlnArgPheG 89
      ::::::::::::::::::::|||
1220 CCAACTACAGCA.....GCCCTAGAGAGCCCGCTGCAGCTGAAG 1263
      ::::::::::::::::::::|||
89 LyGluAspMetArgGlnLeuMetGluLysLeuArgGluArgGlnLeuSer 105
      ::::::::::::::::::::|||
1264 GACAGCAGAGCAGCGTGGCCCGCCAGACCGAGCGGCCCA..... 1307
      ::::::::::::::::::::|||
106 HisSerLeuArgAlaValSerThrAspProHisHisAspHis 121
      ::::::::::::::::::::|||
1308 .....CTACGAGAGCTACAGCACCGCTGAGCCACATCGGCATCAT 1349
      ::::::::::::::::::::|||
```

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-459-504B-6

seq_documentation_block:

; Sequence 6, Application US/08459504B
; Patent No. 6075185

GENERAL INFORMATION:

; APPLICANT: Kozziel, Michael G.
; APPLICANT: Desai, Nalini M.
; APPLICANT: Lewis, Kelly S.
; APPLICANT: Kramer, Vance C.
; APPLICANT: Warren, Gregory W.
; APPLICANT: Evola, Stephen V.
; APPLICANT: Crossland, Lyle D.
; APPLICANT: Wright, Martha S.
; APPLICANT: Merlin, Ellis J.
; APPLICANT: Launis, Karen L.
; APPLICANT: Rothstein, Steven J.
; APPLICANT: Bowman, Cindy G.
; APPLICANT: Dawson, John L.
; APPLICANT: Dunder, Erik M.
; APPLICANT: Pace, Gary M.
; APPLICANT: Suttie, Janet L.
; TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
; TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6075185artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709

COMPUTER READABLE FORM:

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,504B
; FILING DATE:
; CLASSIFICATION:

PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/459,595
; FILING DATE: 02-JUN-1995
; APPLICATION NUMBER: US 07/951,715
; FILING DATE: 25-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/772,027
; FILING DATE: 04-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1577/CIP/DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8587
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 6:

```
SEQUENCE CHARACTERISTICS:
LENGTH: 3624 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3621
OTHER INFORMATION: /product= "Full-length, maize
OTHER INFORMATION: optimized cryiB"
OTHER INFORMATION: /note= "Disclosed in Figure 6."
US-08-459-504B-6

alignment_scores:
Quality: 93.00 Length: 116
Ratio: 1.550 Gaps: 4
Percent Similarity: 51.724 Percent Identity: 26.724

alignment_block:
US-09-327-750D-33 x US-08-459-504B-6
Align seg 1/1 to: US-08-459-504B-6 from: 1 to: 3624
14 GluAsnAspHisGlnLysGluGluLysGluLysProGlnAspH 30
|||||
1026 GAACACGAGCAGCCCGCCAGACACACGATCAACCCGTCGCC 1075
30 rleLysArgGluProValAlaProThrPheGluAlaGlyLysAsnC 47
|||||
1076 TCGCTTCGCCAGCGCGACGCTGTACCGCAGAGAGCTAGCCGCGG 1125
47 ysAla.....ProArgGlyArgArgArg 55
|||||
1126 CTGCTGTGGGCATCTACCTGGAGCCCATCCAGCGTGCACCGTCGC 1175
56 PheArgValArgGlnProIleSerHisTyArgTrpAspLeuMetHisar 72
|||||
1176 CTTCAACTTCAACACCCAGACATACCGA.....CCGGGACCG 1219
72 gValGlyGluProGlnGlyArgMetArgGluGluAsnValGlnArgPheG 89
|||||
1220 CCAACTACAGCA.....GCCCTAGAGAGCCCGCTGCAGCTGAAG 1263
89 LyGluAspMetArgGlnLeuMetGluLysLeuArgGluArgGlnLeuSer 105
|||||
1264 GACAGCAGAGCAGCGTGGCCCGCCAGACCGAGCGGCCCA..... 1307
106 HisSerLeuArgAlaValSerThrAspProHisHisAspHis 121
|||||
1308 .....CTACGAGAGCTACAGCACCGCTGAGCCACATCGGCATCAT 1349
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-459-444-6
```

seq_documentation_block:

; Sequence 6, Application US/08459444A
; Patent No. 6121014

GENERAL INFORMATION:

; APPLICANT: Kozziel, Michael G.
; APPLICANT: Desai, Nalini M.
; APPLICANT: Lewis, Kelly S.
; APPLICANT: Kramer, Vance C.
; APPLICANT: Warren, Gregory W.
; APPLICANT: Evola, Stephen V.
; APPLICANT: Crossland, Lyle D.
; APPLICANT: Wright, Martha S.
; APPLICANT: Merlin, Ellis J.
; APPLICANT: Launis, Karen L.

TITLE OF INVENTION: METHOD FOR PRODUCING A PLANT-OPTIMIZED
NUCLEIC ACID CODING SEQUENCE


```
14 GluAsnAspHisGlnLysLysGluGluLysGluGluLysProGlnAspTh 30
|||||
1026 GAACACGAGCAGCAGCCGCGCCACCAACACGAGCATCAACCCCGTGACCC 1075
30 rIleLysArgGluProValAlaProThrPheGluAlaGlyLysAsnC 47
|||||
1076 TCGCGTTCGCGAGCGGCGAGTGTACCGCAGGAGCTACGCGCGGTG 1125
47 ysAla.....ProArgGlyGlyArgArg 55
|||||
1126 CTGCTGTGGGCACTACCTGGAGCCCATCCAGCGGTGCCACCGCTCG 1175
56 PheArgValArgGlnProIleSerHisTyArgTrpAspLeuMetHisAr 72
|||||
1176 CTCAACTTCAACACCCCGCAGAACATCAGCGA.....CCGCGCAGCG 1219
72 gValGlyGluProGlnGlyArgMetArgGluGluAsnValGlnArgPheG 89
|||||
1220 CCAACTACAGCA.....GCCCTAGCAGAGCCCGCGCTGCAGCTGAAG 1263
89 lyGluAspMetArgGlnLeuMetGluLysLeuArgGluArgGlnLeuSer 105
|||||
1264 GACAGCGAGCGAGCTGCCCCCGAGACCCAGCGCCCA.....1307
106 HisSerLeuArgAlaValSerThrAspProHisHisAspHis 121
|||||
1308 .....CTACGAGAGCTACAGCCAGCGCTGAGCCACATCGGCATCAT 1349
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:us-09-053-549-1
```

```
seq_documentation_block:
; Sequence 1, Application US/09053549
; Patent No. 6121521
; GENERAL INFORMATION:
; APPLICANT: Desai, Nalin
; TITLE OF INVENTION: No. 6121521el Insecticidal Protein and Gene
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6121521artis Corporation
; STREET: 3054 Cornwallis Rd.
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/053,549
; FILING DATE: 01-APR-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: 40,403
; REFERENCE/DOCKET NUMBER: CGC 1995
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8854 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 11..3694
; OTHER INFORMATION: /product- "hyFLIB protein"
```

```
; NAME/KEY: misc_feature
; LOCATION: 8854
; OTHER INFORMATION: /note- "Sequence of pCIB5520
; Patent No. 6121521
; OTHER INFORMATION: containing coding sequence for hyFLIB protein"
; US-09-053-549-1

alignment_scores:
Quality: 91.00 Length: 116
Ratio: 1.542 Gaps: 4
Percent Similarity: 50.862 Percent Identity: 27.586

alignment_block:
US-09-327-750D-3 x US-09-053-549-1
Align seg 1/1 to: US-09-053-549-1 from: 1 to: 8854
14 GluAsnAspHisGlnLysLysGluGluLysGluGluLysProGlnAspTh 30
|||||
1099 GAACACGAGCAGCAGCCGCGCCACCAACACGAGCATCAACCCCGTGACCC 1148
30 rIleLysArgGluProValAlaProThrPheGluAlaGlyLysAsnC 47
|||||
1149 TCGCGTTCGCGTTCGCGAGAGCTGTACCGCAGCAGAGCTACGCGCGGTG 1198
47 ysAla.....ProArgGlyGlyArgArg 55
|||||
1199 CTGCTGTGGGCACTACCTGGAGCCCATCCAGCGGTGCCACCGTGGG 1248
56 PheArgValArgGlnProIleSerHisTyArgTrpAspLeuMetHisAr 72
|||||
1249 CTCAACTTCAACACCCCGCAGAACATCAGCGA.....CCGCGCAGCG 1292
72 gValGlyGluProGlnGlyArgMetArgGluGluAsnValGlnArgPheG 89
|||||
1293 CCAACTACAGCA.....GCCCTAGCAGAGCCCGCGGTTCAGCTGAAG 1336
89 lyGluAspMetArgGlnLeuMetGluLysLeuArgGluArgGlnLeuSer 105
|||||
1337 GACAGCGAGCGAGCTGCCCCCGCAGACCCAGCGCCCA.....1380
106 HisSerLeuArgAlaValSerThrAspProHisHisAspHis 121
|||||
1381 .....CTACGAGAGCTACAGCCAGCGCTGAGCCACATCGGCATCAT 1422
seq_name: /cgn2_5/ptodata/2/ina/6A_COMB.seq:US-08-553C-6

seq_documentation_block:
; Sequence 6, Application US/08506553C
; Patent No. 6120989
; GENERAL INFORMATION:
; APPLICANT: Vornhagen, Rolf; Hinderer, Walter; Schoeneborn, Han-H.;
; APPLICANT: Plachter, Bodo; and Jahn, Gerhard
; TITLE OF INVENTION: ISOLATED HUMAN CYTOMEGALOVIRUS
; TITLE OF INVENTION: POLYPEPTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felte & Lynch
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/506,553C
; FILING DATE: 07/25/95
; CLASSIFICATION: 433
```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary Anne Schofield
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LEDER 202-PFF/MAS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 607 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-08-506-553C-6

alignment_scores:
  Quality: 82.50      Length: 117
  Ratio: 1.650       Gaps: 4
  Percent Similarity: 42.735   Percent Identity: 25.641

alignment_block:
US-09-327-750D-33 x US-08-506-553C-6
..
Align seg 1/1 to: US-08-506-553C-6 from: 1 to: 607

50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyAr 66
|||||
107 CGGTCTGGGAGGAGTGGCGCGTGGCGCAAGAA...GCACGACCGCG 153
66 gTrpAspLeuMethHisArgValGlyGluProGln..... 77
|||||
154 GTGGCGGCGGTGTTCCGGTACGCGGAATGAGTAGCGGTGGCGCGCG 203
77 ..... 77

204 GGTGATCATGACCAGCGTCTTCTCTCAAGAAAAATACGACGACAA 253
78 ..... GlyArgMetArgGlu 83
254 GATCACAGCTACCTGACGTCCTCAAGAGTGATCGCGCGCGGAGGAG 303
83 uAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 100
|||||
304 GAGGAGCGCGGTTTGGATCGCAACTCCGCAATTACTTCAACGACGCG 353
100 rg.....GluArgGlnLeuSer..... 105
354 AAAGAGGAGCGGACGCGGAGGATCTGTACCTTCGATTCGTCCTCAA 403
106 HisSerLeuArgAlaValSerThrAspProProHisHisAspHisAs 122
|||||
404 CACCAAGAAGCAAAAGTGGCGCAAGATCTCTGAAGAGCAGCGCGCATG 453
122 p 122
454 A 454

```

```
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-998-416-994
```

```
seq_documentation_block:
```

```
; Sequence 994, Application US/08998416
```

```
; Patent No. 6239264
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Philippesen, Peter
```

```
; APPLICANT: Pohlmann, Rainer
```

```
; APPLICANT: Steiner, Sabine
```

```
; APPLICANT: Mohr, Christine
```

```
; APPLICANT: Wendland, Jürgen
```

```

; APPLICANT: knechtie, Philipp
; APPLICANT: Rebischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPIT
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 994:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 729 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG16100P
US-08-998-416-994

alignment_scores:
  Quality: 82.00      Length: 112
  Ratio: 1.302       Gaps: 8
  Percent Similarity: 56.250   Percent Identity: 31.250

alignment_block:
US-09-327-750D-33 x US-08-998-416-994/rev
..
Align seg 1/1 to reverse of: US-08-998-416-994 from: 1 to: 729

16 AspHisGlnLysLysGluGluLysGluGluLysProGlnAspThrIle 32
|||||
670 AACACAGAAAGCTCGACAGCAACAGTGATCGCCAGTT..... 632
32 sArgGluProValAlaProThrPheGluAlaGlyLysAsnCysAlap 49
|||||
631 .....GAACCATATATATACCGTTTCCAC.....GCCA 601
49 toArg.GlyGlyArgArgPhe.ArgValArgGlnProIleSerHisT 65
|||||
600 AACCGAGCGACAGAAACCGCTTTGAAATATACGTCACCGAGAGTCAC. 552
65 yTrgTrpAspLeuMethHisArgValGlyGluProGlnGlyArgMetArg 81
|||||
551 ..AGGATGAACGTCGCGCATCTC.....CCCCAG...CGACTGCGT 516
82 GluGluAsnValGlnArgPheGlyGluAspMetArgGlnLeu..... 95
|||||
515 GATCGAGTACACCGCTACCGATGAAGATGTACGCCAACACGACGCCCA 466

```



```

alignment_scores:
  Quality: 80.50      Length: 72
  Ratio: 1.872       Gaps: 1
  Percent Similarity: 59.722   Percent Identity: 31.944

alignment_block:
  US-09-327-750D-33 x US-08-147-777-3
  Align seg 1/1 to: US-08-147-777-3 from: 1 to: 24979

50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyrAr 66
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
5088 CGCGGGAGGGCGGCGAGCTACGCTCAGAGGCTACGCGGCGGCAGCG 5137

66 gTtPAspLeuMethHisArgValGlyGluProGlnGlyArgMetArgGluG 83
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5138 GTGGGGGCTATGGCGCGGAAGAAGCTACCGCGGAGCGGCGGCACGAGGC 5187

83 luAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
5188 GGGAGTGGAGCGGCTTATGCGACGCGCTCGCGCGGCGGTGGCGGAGCTA 5237

100 ...ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPr 115
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
5238 CGGAGGAGCGGCGAGGCGCGGATCCTCGCGCGGTGCAGATCATCC 5287

115 oProHisHisAspHis 120
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5288 AGACCTCCACCAACAC 5303

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-452-872-3

seq_documentation_block:
; Sequence 3..Application US/08452872
; Patent No. 6057298
; GENERAL INFORMATION:
; APPLICANT: Roop, Dennis R.
; APPLICANT: Rothnagel, Joseph A.
; APPLICANT: Greenhalgh, David A.
; APPLICANT: Yuspa, Stuart H.
; TITLE OF INVENTION: KERATIN K1 EXPRESSION VECTORS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LYON & LYON
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452.872
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/147,777
; FILING DATE:
; APPLICATION NUMBER: 07/876,289
; FILING DATE: April 30, 1992
; APPLICATION NUMBER: Unassigned. (204/144)
; FILING DATE: October 29, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 204/153
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24979 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-452-872-3

alignment_scores:
  Quality: 80.50      Length: 72
  Ratio: 1.872       Gaps: 1
  Percent Similarity: 59.722   Percent Identity: 31.944

alignment_block:
  US-09-327-750D-33 x US-08-452-872-3
  Align seg 1/1 to: US-08-452-872-3 from: 1 to: 24979

50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyrAr 66
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
5088 CGCGGGAGGGCGGCGAGCTACGCTCAGAGGCTACGCGGCGGCAGCG 5137

66 gTtPAspLeuMethHisArgValGlyGluProGlnGlyArgMetArgGluG 83
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5138 GTGGGGGCTATGGCGCGGAAGAAGCTACCGCGGAGCGGCGGCACGAGGC 5187

83 luAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
5188 GGGAGTGGAGCGGCTTATGCGACGCGCTCGCGCGGCGGTGGCGGAGCTA 5237

100 ...ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPr 115
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
5238 CGGAGGAGCGGCGAGGCGCGGATCCTCGCGCGGTGCAGATCATCC 5287

115 oProHisHisAspHis 120
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5288 AGACCTCCACCAACAC 5303

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-03985-3

seq_documentation_block:
; Sequence 3..Application PCT/US9303985
; GENERAL INFORMATION:
; APPLICANT: Roop, Dennis R.
; APPLICANT: Rothnagel, Joseph A.
; APPLICANT: Greenhalgh, David A.
; APPLICANT: Yuspa, Stuart H.
; TITLE OF INVENTION: DEVELOPMENT OF A VECTOR TO TARGET GENE
; TITLE OF INVENTION: EXPRESSION TO THE EPIDERMIS OF TRANSGENIC ANIMALS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77010-3095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03985
; FILING DATE: 19930428
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul, Thomas D.

```


Page 11

alignment_scores:		
Quality:	80.50	Length: .72
Ratio:	1.872	Gaps: 1
Percent Similarity:	59.722	Percent Identity: 31.944

alignment_block:

US-09-327-750D-33 x PCT-US93-03985-3

Align seg 1/1 to: PCT-US93-03985-3 from: 1 to: 24979

66 gTrpAspLeuMethHisArgValGlyGluProGlnGlyArgMetArgGluG 83
 1111 111 :::111 111:: 111 :::
 5138 GTGGGGGCTATGGCGCGGAAGAAGCTACCGGGGAGCGGGCGACGAGGC 5187

100 ...ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPr 115
 |||||
 :: ||| :::
 5238 CGGAGGGAGGGCAGACGGCGCGGATCTCTCGCGGCTGCAGATCATCC 5287

115 OPROHISHisASPHis 120

5288 AGACCTCCACCAACAC 5303

248 AGAGAAACCGTAGCGGTTCCGCTTAGCGAGCCATCCCTGCAGTAG 297
 66 gTPAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluG 83
 298 ATGGGACATATCCATAGCTTGGAGAGCCACAGGAGGATGAGAGAGG 347
 83 luAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99
 348 AGAATATGGAAGGATTGGGAGGAGGTGACACAGCTGATGGAAGCTG 397
 100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
 398 AGGAAAGACGATTGAGTCTATCTGCGGAGCTGACACTGATCCGCC 447
 116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128
 448 TCACCATGACCATCAGGATGAGTTTGGCTTATGCC 484

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI58581

seq_documentation_block:

ID AAI58581 standard; CDNA; 862 BP.

XX AC AAI58581;

XX DT 22-OCT-2001 (first entry)

XX DE Human polynucleotide SEQ ID NO 784.

XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-052317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2000; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 29-NOV-2000; 2000US-0727344.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX PI Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX DR WPI: 2001-442253/47.

XX DR P-PSDB; AAM39425.

XX PT Novel nucleic acids and polypeptides, useful for treating disorders

XX PT such as central nervous system injuries -

XX PS Claim 1; SEQ ID NO 784; 10078pp; English.

XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and

XX CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,

XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful

XX CC in gene therapy. A composition containing a polypeptide or polynucleotide

XX CC of the invention may be used to treat diseases of the peripheral nervous

XX CC system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX SQ Sequence 862 BP; 235 A; 183 C; 238 G; 206 T; 0 other;

alignment_scores: 4

Quality: 468.00 Length: 129

Ratio: 4.000 Gaps: 2

Percent Similarity: 90.698 Percent Identity: 68.217

alignment_block:

US-09-327-750D-33 x AAI58581

Align seg 1/1 to: AAI58581 from: 1 to: 862

1 MetGluSerLysAspGlnGlyAla...LysAsnLeuAsnMetGluAsnAs 16

226 ATGGAGTCCAAAGAGGAGGCGGTTAAACAATCTCATCGTGAATAATGT 275

16 pHisGlnLysGluGluLysGluGluLysProGlnAspThrIleLysA 33

276 CAACGAGAAATGATGAAAGATGAAAGGAGGCAAGTTGCTAATAAAG 325

33 rgGluProValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49

326 GGGAGCCCTTGGCCCTACCT...TTCAATGTTAGTGAATACCTGTGCCT 372

50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyxAR 66

373 AGAGAAACCGTAGCGGTTCCGCTTAGCGAGCCATCCCTGCAGTAGTAG 422

66 gTPAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluG 83

423 ATGGACATAATGATAGGCTTGGAGAGCCACAGGCAAGGATGAGAGAGG 472

83 luAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99

473 AGAATATGGAAGGATTGGGAGGAGGTGAGACAGCTGATGAAAGCTG 522

100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116

523 AGGAAAGCAGTTGAGTCATAGTTTTCGGGGCAGTCAGCACTGATCCGCC 572

116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128

573 TCACCATGACCATCAGGATGAGTTTGGCTTATGCC 609

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF59611

seq_documentation_block:

ID AAF59611 standard; CDNA; 898 BP.

XX AC AAF59611;

XX DT 24-APR-2001 (first entry)

XX DE Human cell cycle and proliferation protein CCYPR-22 CDNA, SEQ ID NO:76.

XX KW Cell cycle and proliferation protein; CCYPR; human; agonist;

XX KW antagonist; gene therapy; detection; gene therapy;

XX KW transgenic animal disease model; immune disorder;

XX KW developmental disorder; cell signalling disorder;

XX KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;

XX KW arteriosclerosis; asthma; allergy; diabetes mellitus;

menstrual cycle disorder; bacterial infection; ss.
 OS Homo sapiens.
 XX WO200107471-A2.
 XX PD 01-FEB-2001.
 XX PF 21-JUL-2000; 2000WO-US19948.
 XX PR 21-JUL-1999; 99US-0145075.
 XX PR 08-SEP-1999; 99US-0153129.
 XX PR 10-NOV-1999; 99US-0164647.
 XX PA (INCY-) INCYTE GENOMICS INC.
 XX PI Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman D;
 PI Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;
 XX WPI: 2001-112727/12.
 XX DR P-PSDB; AAB60474.
 XX PT Human cell cycle and proliferation proteins and polynucleotides are
 PT used to treat, diagnose and prevent immune, developmental and cell
 PT signaling disorders and cell proliferative disorders including cancer -
 XX Claim 5; Page 181-182; 205pp; English.
 XX Sequences AAF59590-AAF59643 represent cDNAs encoding 54 human
 CC cell cycle and proliferation proteins (CCYPR), AAB60453-AAB60506.
 CC CCYPR and agonists of CCYPR are used to treat diseases or conditions
 CC associated with decreased expression of functional CCYPR, while CCYPR
 CC antagonists are used to treat diseases or conditions associated with
 CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies
 CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or
 CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect
 CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)
 CC that specifically bind to CCYPR, and in drug screening methods to
 CC identify compounds that modulate the activity of CCYPR. CCYPR
 CC nucleotides can be used to generate transgenic animal models of human
 CC disease, and can be used in gene therapy in target cells with genetic
 CC abnormalities with respect to the expression of CCYPR for the
 CC treatment or prevention of a disorder associated with CCYPR.
 CC Diseases which can be diagnosed, treated and prevented using CCYPR
 CC proteins, nucleic acids, agonists or antagonists include immune,
 CC developmental and cell signalling disorders and cell proliferative
 CC disorders including cancer. Specific examples of these disorders
 CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
 CC diabetes mellitus, disorders of the menstrual cycle and infections
 CC caused by bacteria.
 XX Sequence 898 BP; 250 A; 186 C; 251 G; 211 T; 0 other;

alignment_scores:
 Quality: 458.00 Length: 129
 Ratio: 4.000 Gaps: 2
 Percent Similarity: 90.698 Percent Identity: 68.217

alignment_block:

us-09-327-750d-33 x AAF59611 ..

Align seg 1/1 to: AAF59611 from: 1 to: 898

1 MetGluSerLysAspGlnGlyVala...LysAsnLeuAsnMetGluAsnAs 16
 254 ATGGAGTCCAAAGAGGACGCGTTAAACAATCTCATCTCGTGGAAATGT 303
 16 pHISGlnLysGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 33
 304 CAACACAGGAAATGATGAAAGAGTGAAGAGGACCAAGTTGCTAATTAAG 353
 33 rGluProValValalaProThrPheGluAlaGlyLysAsnCysAlaPro 49

354 GGGAGCCCTTGGCCCTACCT...TTGAATGTTAGTGAATACTGTGTGCT 400
 50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyrAr 66
 401 AGAGAAACCGTAGCGGTTCGGCTTAGGAGCCCATCTCCGACGTATAG 450
 66 gTTPAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluG 83
 451 ATGGGACATAATGCATAGCTTGGAGAGCCACAGCAGGAGGATGAGAGG 500
 83 LuAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99
 501 AGAATATGAAAGGATTGGGAGGAGGTGAGACAGCTGATGGAAAGCTG 550
 100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
 551 AGGAAACAGCTTGAGTCATAGTCTGGGCGAGTCAGCAGCTGATCCCC 600
 116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128
 601 TCACCATGACCATCAGATGAGTTTGGCTTATGCC 637

seq name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH45143

seq_documentation_block:

ID AAH45143 standard; cDNA; 792 BP.

XX AAH45143;

XX AC 07-SEP-2001 (first entry)

XX DE Human brain expressed X-linked protein, hbex, coding sequence.

XX KW Human; brain expressed X-linked protein; cytostatic; auditory; nontropic;
 KW hbex; dysembryoplasia; hereditary disease; cancer; tumour; deafness;
 KW X-chromosome-binding mental retardation; lissencephalous disease; ss.

XX OS Homo sapiens.

XX PN WO200140286-A1.

XX PD 07-JUN-2001.

XX PF 27-NOV-2000; 2000WO-CN00502.

XX PR 30-NOV-1999; 99CN-0124179.

XX PA (BIOR-) BIROAD GENE DEV LTD SHANGHAI.

XX PI Mao Y, Xie Y;

XX DR WPI: 2001-397944/42.

XX DR P-PSDB; AAB99224.

XX PT Isolated human brain-expressed X-linked polypeptide used to diagnose
 PT and treat of dysembryoplasia, hereditary diseases, cancer, tumour,
 PT deafness and X-chromosome-binding mental retardation

XX PS Claim 5; Page 22; 30pp; Chinese.

XX CC The present sequence is the coding sequence for a human brain-expressed
 CC X-linked protein (hbex). hbex and its coding sequence are useful in the
 CC diagnosis and treatment of dysembryoplasia, hereditary diseases, cancer,
 CC tumours, deafness, X-chromosome-binding mental retardation and
 CC lissencephalous disease. hbex is also useful for screening mimics,
 CC agonists, or inhibitors, and in peptide fingerprinting identification.
 CC hbex coding sequence can be used as primers or probes, or in producing
 CC gene chips or microarrays.

XX SQ Sequence 792 BP; 214 A; 172 C; 219 G; 187 T; 0 other;

alignment_scores: Quality: 466.50 Length: 129
 Ratio: 4.092 Gaps: 3
 Percent Similarity: 88.372 Percent Identity: 70.543

alignment_block:
 US-09-327-750D-33 x AAH45143 ..
 Align seg 1/1 to: AAH45143 from: 1 to: 792

1 MetGluSerLysAspGlnGlyAlaLysAsn...LeuAsnMetGluAsnAs 16
 |||||..... ||| ||| |||||.....
 173 ATGGAGTCCAAAGAGAACTAGCAGTAAACAGCTCTCAGCATGGAATAATGC 222

16 PhisGlnLysLysGluGluLysGluLysProGlnAspThrIleLysA 33
 |||||..... |||||.....
 223 CAACCAAGAAATGAAGAAAGAG.....CANGTTGCTAATAAAG 263

33 rGluProValValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49
 |||||..... ||| |||||.....
 264 GGGAGCCCTGGCCCTCCCT...TTGGATGCTGGTGAATACTGTGTGCCT 310

50 ArgGlyGlyArgArgArgPheArgValArgGlnProIleSerHisTyrAr 66
 |||||..... |||||.....
 311 AGAGAAATCGTAGCGGTTCCCGTTAGGCAGCCCATCTCGCATATAG 360

66 gTrpAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluG 83
 |||||..... |||||.....
 361 ATGGATATGATGATAGCTAGCTTGGAGAACACACAGGCAAGGATGAGAGA 410

83 LuAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99
 |||||..... |||||.....
 411 AGAATATGAAAGGATTGGGGGGGTGAGACAGCTGATGGAAGCTG 460

100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
 |||||..... |||||.....
 461 AGGAAAGCAGTGTGACTATGCTCTCGCGGCAGTCAGCACTGACCCCCC 510

116 OHHisAspHisHisAspGluPheCysLeuMetPro 128
 |||||..... |||||.....
 511 TCACCATGACCATCATGATGAGTTTGCCTTATGCCC 547

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT: AAC03880

seq_documentation_block:
 ID AAC03880 standard; cDNA; 662 BP.
 AC AAC03880;
 DT 06-OCT-2000 (first entry)
 DE Human secreted protein 5' EST, SEQ ID NO: 3878.
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 OS Homo sapiens.
 PN EP1033401-A2.
 PD 06-SEP-2000.
 PF 21-FEB-2000; 2000EP-0200610.
 PR 26-FEB-1999; 99US-0122487.
 PA (GEST) GENSET.
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 DR WPI; 2000-500381/45.
 DR P-PSDB; AAG03874.
 XX

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 1: SEQ ID 3878; 71pp + CD-ROM; English.
 XX

CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.
 XX

SQ Sequence 662 BP; 177 A; 144 C; 192 G; 137 T; 12 other;

alignment_scores: Quality: 455.00 Length: 129
 Ratio: 4.027 Gaps: 2
 Percent Similarity: 87.597 Percent Identity: 66.667

alignment_block:
 US-09-327-750D-33 x AAC03880 ..

Align seg 1/1 to: AAC03880 from: 1 to: 662

1 MetGluSerLysAspGlnGlyAla...LysAsnLeuAsnMetGluAsnAs 16
 |||||..... ||| |||||.....
 207 ATGGAGTCCAAAGAGGACGAGCGTTAAACAATCTCATCTGCGAAAATGT 256

16 PhisGlnLysLysGluGluLysGluLysProGlnAspThrIleLysA 33
 |||||..... |||||.....
 257 CAACCAAGAAATGATGAAAGGATGAAAGGAGGACCAAGTGTGCTAATAAAG 306

33 rGluProValValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49
 |||||..... ||| |||||.....
 307 GGGAGCCCTGGCCCTACCT...TTGRATGTYTGTGAATACTGTGTGCCT 353

50 ArgGlyGlyArgArgArgPheArgValArgGlnProIleSerHisTyrAr 66
 |||||..... |||||.....
 354 AGAGAAATCGTAGCGGTTCCCGTTAGGCAGCCCATCTCGCATATAG 403

66 gTrpAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluG 83
 |||||..... |||||.....
 404 ATGGGATATGATGATAGCTTGGAGAACACACAGGCAAGGATGANAGAAG 453

83 LuAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99
 |||||..... |||||.....
 454 AGAATATGAAAGGATTGGGAGGAGGTGAGACAGCTGATGGAAAGCTG 503

100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
 |||||..... |||||.....
 504 AGGAAAGACAGCTTGAGTCATAGTCTCGCGGCAGTCAGCACTGACCCCCC 553

116 OHHisAspHisHisAspGluPheCysLeuMetPro 128
 |||||..... |||||.....
 554 TCACCATGACCATCATGATGAGTTTGCNNWATGCCC 590

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT: AAC03879

seq_documentation_block:
 ID AAC03879 standard; cDNA; 698 BP.
 XX
 AC AAC03879;
 XX

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activity, chemotactic/chemokinetic activity, haemostatic
 CC Activin/inhibin activity, cancer diagnosis and therapy, drug screening,
 CC and thrombolytic activity, arthritis and inflammation, leukaemias and
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 858 BP; 236 A; 175 C; 241 G; 206 T; 0 other;

alignment_scores:
 Quality: 455.00 Length: 130
 Ratio: 3.889 Gaps: 3
 Percent Similarity: 90.000 Percent Identity: 67.692

alignment_block:

US-09-327-750D-33 x AAC160367

Align seg 1/1 to: AAC160367 from: 1 to: 858

1 MetGluSerLysAspGlnGlyAla...LysAsnLeuAsnMetGluAsnAs 16
 215 ATGGAGTCCAAAGAGGACGAGCGTTAAACAATCTCATCTGGAATAATGT 264
 16 phisGlnLysLysGluGlnLysGluGlnLysProGlnAspThrIleLysA 33
 265 CAACGAGGAAATGATGAAAGATGAAAGGAGCAAGTTGCTATAAAG 314
 33 tGluProValValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49
 315 GGGAGCCCTGGCCCTACCT...TTGAATGTTAGTCAATCTGTGCT 361
 50 ArgGlyGlyArgArgArgPheArgValArgGlnProIleSerHisTyra 66
 362 AGAGAAACCCGTAGCGGTTCCGCGTTAGGACGCCCATCTGCAGTATAG 411
 66 gTirAspLeuMethisArgValGlyGluProGlnGlyArgMetArgGluG 83
 412 ATGGACATATATGATAGGCTTGAGAGCCACAGGCAAGGATGAGAGG 461
 83 LuasnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99
 462 AGAATATGGAAGGATGGGAGGAGGTGAGACAGCTGATGGAAGCTG 511
 100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
 512 AGGGAAGAGCAGTTGAGTCAATGCTCGCGGCAGTCAGCACCTGATCCCC 561
 116 ohisHisAspHisAspGluPheCys.LeuMetPro 128
 562 TCACCATGACCATCAGATGATGTTTGCCTTATGCC 599

seq_name: /SIDS2/gcdata/geneseq/geneseq/NA2000.DAT:AAC10889

seq_documentation_block:

ID AAC10889 standard; cDNA; 692 BP.

XX AAC10889;

AC AAC10889;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 14964.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

XX gene therapy; chromosome mapping; ss.

OS Homo sapiens.

XX EP1033401-A2.

XX

PD 06-SEP-2000.
 XX 21-FEB-2000; 2000EP-0200610.
 PF 26-FEB-1999; 99US-0122487.
 PR (GEST) GENSET.
 PA Dumas Milne Edwards J, Duclert A, Giordano J;
 XX WPI; 2000-500381/45.
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX Claim 1; SEQ ID 14964; 71pp + CD-ROM; English.
 XX
 CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 XX
 SQ Sequence 692 BP; 184 A; 151 C; 199 G; 146 T; 12 other;

alignment_scores:
 Quality: 444.50 Length: 125
 Ratio: 4.041 Gaps: 1
 Percent Similarity: 88.000 Percent Identity: 64.800

alignment_block:

US-09-327-750D-33 x AAC10889

Align seg 1/1 to: AAC10889 from: 1 to: 692

4 LysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspHisGlnLysLy 20
 249 AGAGAACGAGCGTTAAACAATCTCATCTGGAATAATGTCACACGAGAAA 298
 20 sGluGluLysGluGlnLysProGlnAspThrIleLysArgGluProVal 37
 299 TGATGAAAGATGAAAGGAGCAAGTGTCTATAAAGGGAGCCCTGG 348
 37 alAlaProThrPheGluAlaGlyLysAsnCysAlaProArgGlyGlyArg 53
 349 CCTACCT...TTGATGTTGTGAATCTCATCTGCTAGAGGAAATCGT 395
 54 ArgArgPheArgValArgGlnProIleSerHisTyraArgTirAspLeuMe 70
 396 AGGCGGTTCCGCGTTAGGACGCCCATCTGCAGTATAGTGGATATGAT 445
 70 thHisArgValGlyGluProGlnGlyArgMetArgGluGluAsnValGlnA 87
 446 GCATAGGCTTGGAGAACACACAGGCAAGGATGANAAGAGAATATGAAA 495
 87 rgPheGlyGluAspMetArgGlnLeuMetGluLysLeuArgGluArgGln 103
 496 GATTGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGGGAAGACAG 545
 104 LeuSerHisSerLeuArgAlaValSerThrAspProProHisHisAspH 120
 546 TTGAGTCATAGTCTCGGGCAGTCAGCACCTGACCCCTCCATCATGACCA 595

120 shisaspGluPheCysLeuMetPro 128
 |||||
 596 TCATGATGAGTTTGCNNWATGCC 620

seq_name: /SIDS2/gcgdata/geneseq/NA2001.DAT:AAF23529

seq_documentation_block:

ID AAF23529 standard; DNA; 891 BP.

XX AAF23529;

XX 22-MAR-2001 (first entry)

XX Human NADE DNA.

XX Neurotrophin receptor; p75-NTR; NGF-induced apoptosis;
 KW neurogenetic disease; NF-kappaB; ds.

XX Homo sapiens.

XX WO200075278-A2.

XX 14-DEC-2000.

XX 07-JUN-2000; 2000WO-US15621.

XX 07-JUN-1999; 99US-0327750.

XX (UYCO) UNIV COLUMBIA NEW YORK.

XX Sato T;

XX WPI; 2001-061707/07.

XX New p75-neurotrophin receptor-associated cell death executor (NADE) and
 PT the gene encoding NADE, useful for modulating the activity of p75NTR
 PT and for detecting neurodegenerative diseases -

XX Disclosure; Fig 1; 134pp; English.

XX The present invention relates to a purified polypeptide capable of
 CC binding neurotrophin receptor (p75-NTR). The invention is useful for
 CC binding and modulating the activity of p75NTR. The peptide mediates
 CC NGF-induced apoptosis, which plays an important role in neurogenetic
 CC diseases. The peptide of the invention and p75NTR are useful for
 CC inhibiting NF-kappaB activation in a cell or a subject, for inducing
 CC caspase-2 and caspase-3 activity to cleave poly (ADP-ribose) polymerase
 CC and fragment nuclear DNA in a cell by co-expression of (I) and p75-NTR.

XX Sequence 891 BP; 251 A; 182 C; 224 G; 234 T; 0 other;

alignment_scores:

Quality: 187.50 Length: 129
 Ratio: 2.232 Gaps: 5
 Percent Similarity: 65.116 Percent Identity: 36.434

alignment_block:

US-09-327-750d-33 x AAF23529 ..

Align seq 1/1 to: AAF23529 from: 1 to: 891

2 GluSerLysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspHisG1 18

279 AAGAAACACACACACCAAAAAAATCTCATCATGCGCAATATTCACCA 328

18 nLysLysGluGluLysProGlnAspThrIleLysArgGluP 35

329 GGAACACGAGAGATGGAGCGCTATGCAGATGGAGGAGACCGCC 378

35 roValValaProThrPheGluAlaGlyLysAsnCysAlaProArgGly 51

379 CTTTG.....GGAGAGGTGAAGCCACCAGCCT...GCA 410
 52 GlyArgArgPheArgValArgGlnProIleSerHisTyrArgTrpAs 68
 |||||
 411 GGAATTCGACGGGACAGGCTCGCGACTTCCCTTAATTTTCGATGGC 460
 68 pLeuMetHisArg...ValGlyGluProGlnGlyArgMetArgGluGluA 84
 :|||
 461 CATACCAATAGCAGATCAATGATGGATGGTGA...GATGGAGATG 507
 84 snValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuArg 100
 :|||
 508 ATATGCAATATTTCATGGAGGAGATGAGAGAAATCAGAGAAAACCTAGG 557
 101 GluArgGluLeuSerHisSerLeuArgAlaValSerThrAsp...ProPr 116
 |||||
 558 GAGCTGCAGTTGAGGAATTGCTGCGTATCCTTATGGGGAGGCTCTCTAA 607
 116 ohHisAspHisHisAspGluPheCysLeuMetPro 128
 |||||
 608 TCACCATGACCATCATGATGAATTTTGCCTTATGCCT 644

seq_name: /SIDS2/gcgdata/geneseq/NA2001.DAT:AAF23528

seq_documentation_block:

ID AAF23528 standard; DNA; 700 BP.

XX AAF23528;

XX 22-MAR-2001 (first entry)

XX Mouse NADE DNA.

XX Neurotrophin receptor; p75-NTR; NGF-induced apoptosis;
 KW neurogenetic disease; NF-kappaB; ds.

XX Mus sp.

XX WO200075278-A2.

XX 14-DEC-2000.

XX 07-JUN-2000; 2000WO-US15621.

XX 07-JUN-1999; 99US-0327750.

XX (UYCO) UNIV COLUMBIA NEW YORK.

XX Sato T;

XX WPI; 2001-061707/07.

XX New p75-neurotrophin receptor-associated cell death executor (NADE) and
 PT the gene encoding NADE, useful for modulating the activity of p75NTR
 PT and for detecting neurodegenerative diseases -

XX Claim 12; Fig 1; 134pp; English.

XX The present invention relates to a purified polypeptide capable of
 CC binding neurotrophin receptor (p75-NTR). The invention is useful for
 CC binding and modulating the activity of p75NTR. The peptide mediates
 CC NGF-induced apoptosis, which plays an important role in neurogenetic
 CC diseases. The peptide of the invention and p75NTR are useful for
 CC inhibiting NF-kappaB activation in a cell or a subject, for inducing
 CC caspase-2 and caspase-3 activity to cleave poly (ADP-ribose) polymerase
 CC and fragment nuclear DNA in a cell by co-expression of (I) and p75-NTR.

XX Sequence 700 BP; 177 A; 188 C; 203 G; 132 T; 0 other;

alignment_scores:

Quality: 173.00 Length: 139
 Ratio: 2.190 Gaps: 5

Percent Similarity: 56.835 Percent Identity: 32.374

alignment_block:

US-09-327-750D-33 x AAF23528

Align seg 1/1 to: AAF23528 from: 1 to: 700

```
9  LysAsnLeuAsnMetGluAsnAspHisGlnLysLysGluGluLysGluGlu 25
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
165 CAAATCTCATCTGCGGCAATGTCCACGAGAAAGAGAGCTGGAGCA 214
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
25  uLysProGlnAspThrIleLysArgGluProValValAlaProThrPheG 42
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
215 GCCCTGCGAGATGACGAGAGACGCCCTGTG.....G 249
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
42  luAlaGlyLysAsnGlyAlaProArgGlyGly..... 52
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
250 GAGGAGGTGAGGGCCACCGCTGCTGCAACAAACAAACAAACACAC 299
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
53  .....ArgArgArgPheArgValArgGlnProIle 63
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
300 AACCATACCAACACACCCACCGAGAGAGCCAGGCTGCGGACTGCC 349
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
63  rHisTyrArgTrpAspLeuMetHisArgValGlyGluProGlnGlyArgM 80
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
350 TAACTTCGATGGGCCATT.....CCCAACAGGCAGA 381
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
80  eArgGlu.....GluAsnValGlnArgPheGlyGlu 90
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
382 TGAATGACGGGTGGTGGAGATGATGATGATGATGATGATGATGATG 431
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
91  AspMetArgGlnLeuMetGluLysLeuArgGluGlnLeuSerHis 107
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
432 GAGATGAGAGATCCGAGAGAACTTAGGGAGCTACAGCTGAGAAATG 481
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
107 rleuArgAlaValSerThrAsp...ProHisHisAspHisHisAspG 123
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
482 TCTACGCATCTTTGGGGGAGGTGCTAACCCACGATCACCATGATG 531
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
123 luPheCysLeuMetPro 128
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
532 AATTCTGCTTATGCT 548
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH03517

seq_documentation_block:

ID: AAH03517 standard; cDNA; 865 BP.

XX AC AAH03517;

XX DT 26-JUN-2001 (first entry)

XX DE Human cDNA clone (5'-primer) SEQ ID NO:352.

XX KW Human; primer; detection: diagnosis; antisense therapy; gene therapy; ss.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
full-length cDNAs defined in the specification, and for the detection
and/or diagnosis of the abnormality of the proteins encoded by the
full-length cDNAs -

PS Claim 1: SEQ ID 352: 2537pp + CD ROM; English.

XX CC The present invention describes primer sets for synthesizing 5602
full-length cDNAs defined in the specification. Where a primer set
comprises: (a) an oligo-dT primer and an oligonucleotide complementary
to the complementary strand of a polynucleotide which comprises one of
the 5602 nucleotide sequences defined in the specification, where the
oligonucleotide comprises at least 15 nucleotides; or (b) a combination
of an oligonucleotide comprising a sequence complementary to the
complementary strand of a polynucleotide which comprises a 5'-end
sequence and an oligonucleotide comprising a sequence complementary to a
polynucleotide which comprises a 3'-end sequence, where the
oligonucleotide comprises at least 15 nucleotides and the combination of
the 5'-end sequence/3'-end sequence is selected from those defined in
the specification. The primer sets can be used in antisense therapy and
in gene therapy. The primers are useful for synthesizing polynucleotides,
particularly full-length cDNAs. The primers are also useful for the
detection and/or diagnosis of the abnormality of the proteins encoded by
cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
AAH13633 to AAH18742 represent human cDNA sequences; AAH2446 to
AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
represent oligonucleotides, all of which are used in the exemplification
of the present invention.

XX SQ Sequence 865 BP; 239 A; 177 C; 224 G; 221 T; 4 other;

alignment_scores:

Quality: 170.00 Length: 130

Ratio: 1.954 Gaps: 5

Percent Similarity: 66.923 Percent Identity: 33.077

alignment_block:

US-09-327-750D-33 x AAH03517

Align seg 1/1 to: AAH03517 from: 1 to: 865

```
1  MetGluSerLysAspGln...GlyAlaLysAsnLeuAsnMetGluAsnAs 16
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
191 ATGGAGTCCAAAGAGAACTAGCGGCAACAATCTCAACGGGAAATGC 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
16  phisGlnLysLysGluGluLysGluLysProGlnAspThrIleLysA 33
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
241 CCAACAAGAAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 290
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
33  rgGluProValValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
291 AATCCCGCCATTGGGA.....GGGGGTGAAGGCCAGAGCCT 328
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
50  ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyrAr 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
329 GGAGGAATATACGCGGGGGGAGTAGGGGAGCTTGTCCCTAATTTTCG 378
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
66  gTrpAspLeuMetHisArg...ValGlyGluProGlnGlyArgMetArg 82
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
379 ATGGGCCATACCTAATAGCATATTGAGCACAATGAAGCGAGA..... 421
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
82  luGluAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLys 98
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
422 ..GATGATGAGAAAGGTTTGTAGGCAGATGATGGAATCAAGAGAAAG 469
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
99  LeuArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPr 115
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


PA (INCY-) INCYTE GENOMICS INC.
 XX Tang YT, Hillman JL, Yue H, Reddy R, Lal P, Shah P, Azimzai Y;
 PI Baughn MR, Lu DAM, Bandman O, Shih LL, Patterson C;
 XX WPI; 2001-211447/21.
 DR P-PSDB; AAB47126.
 XX
 XX Isolated polypeptides and polynucleotides involved in cell
 PT differentiation are used for treatment, prevention and diagnosis of
 PT cell proliferative, developmental and neurological disorders e.g.
 PT cancer and Alzheimer's disease
 XX
 PS Claim 5; Page 121; 137pp; English.
 XX The sequences given in AAC85545-72 encode human polypeptides¹ involved
 CC in cell differentiation (CDIFF). CDIFF polypeptides and agonists of
 CC these are used to treat a disease or condition associated with
 CC decreased expression of functional CDIFF. An antagonist of CDIFF is
 CC used to treat a disease or condition associated with over expression
 CC of functional CDIFF. CDIFF polypeptides may be used for the treatment,
 CC prevention and diagnosis of cell proliferative, developmental and
 CC neurological disorders, such as Alzheimer's disease, schizophrenic
 CC disorders, arteriosclerosis, cancer, atherosclerosis, diabetes mellitus
 CC and epilepsy. The CDIFF-4 sequence is homologous to Mus musculus
 CC REX-3. This sequence maps to chromosome 1 within the interval from
 CC 152.2 to 157.4 centimorgans, to chromosome 3 within the interval from
 CC 157.4 to 158.0 centimorgans, and to the X chromosome within the interval
 CC from 104.9 to 150.3 centimorgans.
 XX
 SQ Sequence 1364 BP; 411 A; 269 C; 322 G; 362 T; 0 other;

alignment_scores:
 Quality: 170.00 Length: 130
 Ratio: 1.954 Gaps: 5
 Percent Similarity: 66.923 Percent Identity: 33.077

alignment_block:
 US-09-327-750D-33 x AAC85548 ..
 Align seg 1/1 to: AAC85548 from: 1 to: 1364

```

1 MetGluSerLysAspGln...GlyAlaLysAsnLeuAsnMetGluAsnAs 16
|||||
250 ATGGAGTCCAAAGAGGAACTAGCGGCAACAACTCTCAGCGGGAATGCC 299
16 pHisGlnLysLysGluGluLysGluGluLysProGlnAspThrIleLysA 33
|||||
300 CCAACAAGAAACGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 349
33 rGluProValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49
|||||
350 ATCCCGCCATTTGGGA.....GGGGTGAAGCCGAGCGCT 387
50 ArGGlyGlyArgArgPheArgValArgGlnProIleSerHisTyrAr 66
|||||
388 GGAGGAATATCAGCGGGGCGAGTGTAGCGACTTGTCCCTAATTTTCG 437
66 qTrpAspLeuMetHisArg...ValGlyGluProGlnGlyArgMetArg 82
|||||
438 ATGGCCATCTATAGGCAATATGAGCAATGAGCGAGA..... 480
82 LuGluAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLys 98
|||||
481 ..GATGATGAGAAAGTTTCTAGCGCAGATGATGGAATCAAGAGAAAG 528
99 LeuArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPr 115
|||||
529 ACTAGGAGCAACGATGAGGACCTATATCGCTTCCAAACCTCTGAACC 578
115 oProHisHisAspHisHisAspGluPheCysLeuMetPro 128
|||||

```

579 T.....GACAACCATATGACTTTTGCTCATACCT 609
 seq_name: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT:AAF21748
 seq_documentation_block:
 ID AAF21748 standard; DNA; 917 BP.
 XX
 AC AAF21748;
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE Human breast and ovarian cancer associated antigen gene SEQ ID 135.
 XX
 KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KW neutropenic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antinflammatory; antitumor; anticonvulsant;
 KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO2000055173-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05881.
 XX
 PR 12-MAR-1999; 9905-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Ruben SM;
 XX
 DR WPI: 2000-611515/58.
 DR P-PSDB; AAB58845.
 XX
 PT New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases .
 XX
 Claim 1; Page 581-582; 1299pp; English.
 XX
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC neutropenic; neuroprotective; antiviral; antiallergic; hepatotropic;
 CC antidiabetic; antinflammatory; antitumor; anticonvulsant;
 CC antibacterial; antifungal; antiparasitic and cardiant activity. The
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and antagonists may also be used in the diagnosis,
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemia; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.
 XX
 SQ Sequence 917 BP; 228 A; 203 C; 270 G; 211 T; 5 other;

alignment_scores:
 Quality: 158.50 Length: 113
 Ratio: 2.142 Gaps: 5
 Percent Similarity: 65.487 Percent Identity: 35.398

```

alignment_block:
US-09-327-750D-33 x AAF21748
Align seg 1/1 to: AAF21748 from: 1 to: 917
18 GlnLysLysGluGluLysProGlnAspThrIleLysArgG1 34
   |||:||||| |||:| |||:| |||:| |||:|
306 CAGGAAACGAGAGATGGAGCAGCCTATCGAAGTGGAGAACG 355
   |||:||||| |||:| |||:| |||:| |||:|
34 uProValValalaProThrPheGluAlaGlyLysAsnCysAlaProArg 51
   |||:||||| |||:| |||:| |||:| |||:|
356 CCCTTTG.....GGAGGAGTGAAGGCCACCGCT...G 387
51 LyGlyArgArgPheArgValArgGlnProIleSerHisTyrArgTtp 67
   |||:||||| |||:| |||:| |||:| |||:|
388 CAGGAAATCGAGGGGACAGCGCTCGCGACTTCCCTTAATTTTCATGG 437
   |||:||||| |||:| |||:| |||:| |||:|
68 AspLeuMetHisArg...ValGlyGluProGlnGlyArgMetArgGlu 83
   |||:||||| |||:| |||:| |||:| |||:|
438 GCCATACCAATAGGCAGATCAATGATGGATGGTGA...GATGGAGA 484
83 uAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuA 100
   |||:||||| |||:| |||:| |||:| |||:|
485 TCATATGGAATATTCATGGAGGAGATGAGAAATCAGAGAAACTTA 534
100 rgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAsp...Pro 115
   |||:||||| |||:| |||:| |||:| |||:|
535 GGGAGCTGCAGTTGAGGAATTTGCTGCGTATCCTTATGGGGAGCTCTCT 584
116 ProHisHisAspHisHisAspGluPheCysLeuMetPro 128
   |||:||||| |||:| |||:| |||:| |||:|
585 AATCACCATGACATCATGATGATTTTGCTTATGCTTATGCTT 623

seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA2000.DAT:AAAC01005
seq_documentation_block:
ID AAC01005 standard; cDNA; 532 BP.
XX
XX AAC01005;
XX
XX 06-OCT-2000 (first entry)
XX Human secreted protein 5' EST, SEQ ID NO: 1003.
XX Human, 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX Homo sapiens.
XX
XX EPI033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI: 2000-500381/45.
XX P-PSDB; AAG00999.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 1003; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. An ORF has been identified within the

```

```

CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
XX Sequence 532 BP; 151 A; 121 C; 163 G; 91 T; 6 other;
SQ
alignment_scores:
Quality: 135.50 Length: 120
Ratio: 1.831 Gaps: 5
Percent Similarity: 61.667 Percent Identity: 32.500
alignment_block:
US-09-327-750D-33 x AAC01005
Align seg 1/1 to: AAC01005 from: 1 to: 532
4 LysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspHisGlnLysLy 20
   |||:| |||:| |||:| |||:| |||:|
192 AAACACCCAGAAAAAAATCTCATCTATGCAATATTCACCAAGAAA 241
20 sGluGluLysGluGluLysProGlnAspThrIleLysArgGluProVal 37
   |||:| |||:| |||:| |||:| |||:|
242 CGAAGAGATGGAGCAGCCTATCCAGATGGAGAGAGACCGCCCTTG. 290
37 alaProThrPheGluAlaGlyLysAsnCysAlaProArgGlyGlyArg 53
   |||:| |||:| |||:| |||:| |||:|
291 .....GGAGGAGTGAAGGCCACCGCT...GCAGGAAT 323
54 ArgArgPheArgValArgGlnProIleSerHisTyrArgTtpAspLeu 70
   |||:| |||:| |||:| |||:| |||:|
324 CGACGGGAGAGGCTCGCGAYTTGCCCTAATTTTCATGGCCATACC 373
70 tHisArg...ValGlyGluProGlnGlyArgMetArgGluAsnValG 86
   |||:| |||:| |||:| |||:| |||:|
374 CAATAGGCAGATCAATGATGGATGGGT...GRAGATGGAGATGATATGG 420
86 InArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuArgGlu 102
   |||:| |||:| |||:| |||:| |||:|
421 AATATTCATGGAGAGATGAGAAATCAGAAACAACTTAGGAGCTG 470
103 GlnLeuSerHisSerLeuArg...AlaValSerThrAspProHisH 118
   |||:| |||:| |||:| |||:| |||:|
471 CAKTTGAGGAATTTGCTGCTATCTTATGGGGAKTCTCTTAATCACCA 520
118 sAspHisHis 121
   |||:| |||:| |||:| |||:| |||:|
521 TGACCATCAT 530

```

OM of: US-09-327-750D-33 to: GenEmbl.* out_format : pfs
 Date: Mar 11, 2002 3:34 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL-frame*_bzn.model -DEV-xlh
-o/cgn2_l/USPOT_spool/US09327750/runat_11032002_101154_20319/app_query.fasta_1.1472
-b/GenEmbl -QFMT-fastap -SUFFIX-rge -GAPOP-12.000 -GAPEXT-4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-OCAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -GAPOP=6.000
-FCAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MAPFIX-biosum62 -TRANS-human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODEL-LOCAL -OUTFMT=pfs -NORM-ext -HEAPSIZ=500
-MINLEN=0 -MAXLEN=200000000 -USER-US09327750 -CGN1_1-8673
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1
```

Search information block:

Query: US-09-327-750D-33
 Query length: 128
 Database: GenEmbl.*
 Database sequences: 1472140
 Database length: 341344837
 Search time (sec): 4557.230000

score_list:

Sequence	Strd Orig	zScore	EScore	Len	Description
gb.ro:AF097437	+	646.00	1.1e-46	2269	1 AF097437 Mus musculus Bexl protein
gb.ro:AF097438	+	640.00	986.65	835	1 AF097438 Mus musculus brain exp
gb.ro:AF051347	+	640.00	984.99	838	1 AF051347 Mus musculus REX-3 mRN
gb.ro:AF097439	+	601.50	926.17	785	1 AF097439 Mus musculus brain exp
gb.ro:BC003254	+	582.00	866.16	2901	1 BC003254 Mus musculus, Similar
gb.pr:AF183416	+	471.50	727.38	642	1 AF183416 Homo sapiens ovarian
gb.pr:AF220189	+	471.50	725.44	828	1 AF220189 Homo sapiens uncharac
gb.pr:HS198P4	+	468.00	720.87	35714	1 AF208708 Human DNA sequence
gb.pr:AF251053	+	468.00	719.43	744	1 AF251053 Homo sapiens X-linked
gb.pr:AF078272	+	468.00	692.37	898	1 AF078272 Sequence 76 from Paten
gb.pr:HSV870H8	+	468.00	690.40	40584	1 AL133348 Human DNA sequence
gb.pr:AL133348	+	468.00	690.32	41029	1 AL022169 Homo sapiens chromos
gb.pr:AF237783	+	466.50	718.09	791	1 AF237783 Homo sapiens brain-exp
gb.sts:G24641	+	311.00	481.91	504	1 G24641 human STS WI-11354, sequ
gb.pr:HSV251F8	+	195.50	265.60	45678	1 270719 Human DNA sequence
gb.pr:HSV251F8	+	187.50	287.27	891	1 M38188 Human unknown protein fr
gb.pr:AF187084	+	187.50	287.27	891	1 AF187084 Homo sapiens p75NTR-as
gb.pr:BC003190	+	186.50	286.62	793	1 BC003190 Homo sapiens, p75NTR-as
gb.ro:AF187086	+	174.00	268.30	700	1 AF187086 Mus musculus p75NTR-as
gb.ro:AF097440	+	174.00	266.79	854	1 AF097440 Mus musculus brain exp
gb.sts:G35294	+	170.00	265.06	477	1 G35294 human STS SHGC-37409, se
gb.pr:AK000959	+	170.00	257.85	1229	1 AK000959 Homo sapiens cDNA FLJ
gb.pr:AK000959	+	170.00	257.06	1364	1 AX100231 Sequence 32 from Paten
gb.pr:HS635G19	+	170.00	227.10	69648	1 AL035494 Human DNA sequence
gb.ro:AF187085	+	148.00	230.52	519	1 AF187085 Rattus norvegicus p75N
gb.sts:G72708	+	142.50	223.64	421	1 G72708 MARC 4953-4954:991939031
gb.ba:SC641	+	102.00	127.34	36028	1 AL442120 Streptomyces coelic
gb.sts:G23964	+	100.50	159.86	372	1 G23964 human STS WI-15922, sequ
gb.pr:AB040971	+	97.50	136.41	4408	1 AB040971 Homo sapiens mRNA for
gb.hlg:AC021617	+	94.50	103.23	187075	1 AC021617 Homo sapiens clone
gb.pat:AF25464	+	93.50	142.84	844	1 A25464 B.pertussis STX pilin ge
gb.ba:BBF1MX	+	93.50	142.83	1115	1 Y00556 Bordetella pertussis fim
gb.pr:AF187085	+	93.50	139.46	1718	1 X74118 B.bronchiseptica fimx g
gb.pat:AR029509	+	93.00	130.97	3624	1 AR029509 Sequence 1 from Patent
gb.pat:AR098462	+	93.00	130.97	3624	1 AR098462 Sequence 6 from Patent
gb.pat:141422	+	93.00	130.97	3624	1 I41422 Sequence 6 from Patent
gb.pr:WACP53A	+	92.00	133.28	2184	1 L20442 Rhesus monkey p53 mRNA,
gb.ba:HV095372	+	92.00	122.97	8455	1 U95372 Haloflex volcanii plas

```
seq_name: gb_ro:AF097437
seq_documentation_block:
LOCUS AF097437 2269 bp DNA ROD 05-AUG-1999
DEFINITION Mus musculus Bexl protein (Bexl) gene, complete cds.
ACCESSION AF097437
VERSION AF097437.1 GI:5702155
KEYWORDS house mouse,
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2269)
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Bexl, a gene with increased expression in parthenogenetic embryos,
is a member of a novel gene family on the mouse X chromosome
JOURNAL Hum. Mol. Genet. 8 (4), 611-619 (1999)
MEDLINE 99172070
PUBMED 10072429
REFERENCE 2 (bases 1 to 2269)
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1998) Cancer Unit, Queensland Institute of
Medical Research, Herston Rd, Brisbane, Qld 4030, Australia
FEATURES
source
1..2269
/organism="Mus musculus"
/strain="129SVJ"
/db_xref="taxon:10090"
/chromosome="X"
/map="near plp"
/tissue_type="liver"
join(409..509,866..942,1263..1873)
/product="Bexl protein"
/product="Bexl protein"
/gene="Bexl"
/notes="expressed in brain; X-linked"
1268..1654
/gene="Bexl"
/codon_start=1
/product="Bexl protein"
/protein_id="AA047168.1"
/db_xref="GI:5702156"
BASE COUNT 563 a 554 c 705 g 447 t
ORIGIN
1
Quality: 646.00 Length: 128
Ratio: 5.168 Gaps: 0
Percent Similarity: 97.656 Percent Identity: 92.969
alignment_block:
US-09-327-750D-33 x AF097437
Align seg 1/1 to: AF097437 from: 1 to: 2269
1 MetGluSerLysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspH1 17
|||||
1268 ATGGAGTCCAAAGATCAAGCGCTGAAAAATCTCAACATGGAGATGACCA 1317
17 scintLysLysGluGluLysGluGluLysProGlnAspThrLysArgG 34
|||||
```


Erratum: [[published erratum appears in Hum Mol Genet 1999 May;8(5):943]]
2 (bases 1 to 785)
Brown, A.L. and Kay, G.F.
Direct Submission
Submitted (08-OCT-1998) Cancer Unit, Queensland Institute of Medical Research, Herston Rd, Brisbane, Qld 4029, Australia
Location/Qualifiers
Source
1..785
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="dbEST:W48832"
/chromosome="X"
/map="near plp"
/dev_stage="embryo; 15.5 dpc"
1..785
/gene="Bex2"
/gene="Bex2"
/codon_start=1
/product="brain expressed x-linked protein 2"
/protein_id="AAD24430.1"
/db_xref="GI:4580592"
/translation="MESKYEQGVKLNNDHOKKEKEKPODASKRDPITVALPFEA
QYIVPRGRRRRFRQPIVHYRNDLMHVRVGEPOGRMEENVORFGDDVRLMEKLE
RQLSHSLRAVSTDPHHDHDEFCLMP"
BASE COUNT 212 a 163 c 235 g 175 t
ORIGIN

alignment_scores:
Quality: 601.50 Length: 129
Ratio: 4.971 Gaps: 1
Percent Similarity: 93.798 Percent Identity: 86.822
alignment_block:
US-09-327-750D-33 x AF097439 ..
Align seg 1/1 to: AF097439 from: 1 to: 785

1 MetGluSerLys...AspGlnGlyAlaLysAsnLeuAsnMetGluAsnAs 16
139 ATGGAGTCCAAAGTGGCAAGCGCTGAAATCTCAACATGGAGATGA 188
16 pHisGlnLysLysGluGluLysGluGluLysProGlnAspThrIleLysA 33
189 CCATCAGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 238
33 rGluProValValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49
239 GGGATCCGATTTGGCCCTGCTTCGAGCTGGAGACTACTACGTCCT 288
50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyrAr 66
289 AGAGGAGGTCCAGCGGTTCGGGTTCGGAGCCCATCGTCACACAG 338
66 gTrpAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgLug 83
339 ATGGGACCTGATGCATAGGTTGGGGAGCCCGGAGGAGGAGGAGGAGG 388
83 luAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99
389 AGAACGTACAGAGTTGGGGATGATGTGACACAGCTCATGTGAGAACTG 438
100 ArgGluArgGlnLeuSerHisSerHisLeuArgAlaValSerThrAspPro 116
439 AGGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 488
116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128
489 TCATCATGACCACTGATGATGATGATGATGATGATGATGATGATGATG 525

/note="retinoic acid reduced gene expression"
/codon_start=1
/product="BEX-3"
/protein_id="AAC61929.1"
/db_xref="GI:3510643"
/translation="MENDHOKKEKEKPODITRPAVALISEAGKNCAPRGRRRR
RVQPIAHYRNDLMQVRVGEPOGRMEENVORFGDDVRLMEKLEKRLSHSLRAVST
DPHHDHDEFCLMP"
BASE COUNT 238 a 181 c 251 g 168 t
ORIGIN

alignment_scores:
Quality: 640.00 Length: 128
Ratio: 5.161 Gaps: 0
Percent Similarity: 96.875 Percent Identity: 92.188
alignment_block:
US-09-327-750D-33 x AF051347 ..
Align seg 1/1 to: AF051347 from: 1 to: 838

1 MetGluSerLysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspHi 17
193 ATGGAGTCCAAAGTCAAGCGCTGAAATCTCAACATGGAGATGACCA 242
17 sGlnLysLysGluGluLysGluGluLysProGlnAspThrIleLysArg 34
243 TCAGAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 292
34 luProValValAlaProThrPheGluAlaGlyLysAsnCysAlaProArg 50
293 AGCCAGCTGTGGCCCTGATCTCCAGGCTGGCAAAACCTGCGCCCTAGA 342
51 GlyGlyArgArgPheArgValArgGlnProIleSerHisTyrArgTr 67
343 GGAGGTCGACGCGGTTCGGGTTCGGCAGCCCATCGCTCATATAGATG 392
67 pAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluLys 84
393 GGACCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 442
84 snValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuArg 100
443 ACCTACAGAGTTGGGGGTGATGTGAGACAGCTCATGGAGAGCTGAGG 492
101 GluArgGlnLeuSerHisSerHisLeuArgAlaValSerThrAspProHi 117
493 GAAGGAGCTGAGCCAGCTCGGGCGGTAGCACCTGACCTGACCCGCTCA 542
117 shisAspHisHisAspGluPheCysLeuMetPro 128
543 TCATGACCACTGATGATGATGATGATGATGATGATGATGATGATG 576

seq_name: gb_ro:AF097439
seq_documentation_block:
LOCUS AF097439 785 bp mRNA 13-APR-1999
DEFINITION Mus musculus brain expressed x-linked protein 2 (Bex2) mRNA,
complete cds.
ACCESSION AF097439
VERSION AF097439.1 GI:4580591
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 785)
AUTHORS Brown, A.L. and Kay, G.F.
TITLE Bex1, a gene with increased expression in parthenogenetic embryos,
is a member of a novel gene family on the mouse X chromosome
JOURNAL Hum. Mol. Genet. 8 (4), 611-619 (1999)
MEDLINE 99172070

/translation="MESKEKRAVNSLSMENAOENKEEOVANKGEPALPLDAGEYC
VPRGNRRFRVQPILQYRWDMHRLGEPQARRENNRIGEEVRLMEKLEKREKOLS
HSLRAVSTDPHHDHDFCLMP"

BASE COUNT 204 a 118 c 157 g 163 t
ORIGIN

alignment_scores:
Quality: 471.50 Length: 129
Ratio: 4.100 Gaps: 3
Percent Similarity: 89.147 Percent Identity: 70.543

alignment_block:

US-09-327-750D-33 x AF183416

Align seg 1/1 to: AF183416 from: 1 to: 642

```

1 MetGluSerLysAspGlnGlyAlaLysAsn...LeuAsnMetGluAsnAs 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3 ATGGAGTCCAAAGAGAACAGCAGTAAACAGTCTCAGCATGGAATAATGC 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
16 phisGlnLysLysGluLysGluLysProGlnAspThrIleLysA 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
53 CAACCAAGAAATGAAGAAAGAG.....CAAGTTGCTAATAAAG 93
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
33 rgGluProValValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
94 GGGAGCCCTTGGCCCTCCCT...TTGGATGCTGGTGAATCTGTGCT 140
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyrAr 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
141 AGAGAAATCTAGCGGTCCCGTTAGGCACCCCATCTCGCAGTATAG 190
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
66 gtrpAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGlu 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
191 ATGGGATATGATGATAGGTGGAGAACACACAGCAAGATGAGAGAAG 240
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
83 LuAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 AGAATATGGAAGGATGGGAGGAGGTGAGACAGCTGATGGAAGCTG 290
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
291 AGGGAAGAGCAGTGTAGTCTAGTCTCGGGCAGTCAAGCACTGACCC 340
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116 ohisHisAspHisAspGluPheCysLeuMetPro 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
341 TCACCATGACCATCATGATGAGTTTTCCTTATGCC 377
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: gb_pr:AF220189

seq_documentation_block:

LOCUS AF220189 828 bp mRNA PRI 04-MAY-2000
DEFINITION Homo sapiens uncharacterized hypothalamus protein HBEX2 mRNA,
complete cds.

ACCESSION AF220189

VERSION AF220189.1 GI:7689028

KEYWORDS human.

SOURCE

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Xiao.H., Song.H., Gao.G., Ren.S., Chen.Z. and Han.Z.

TITLE A novel gene expressed in human hypothalamus

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 828)

AUTHORS Xiao.H., Song.H., Gao.G., Ren.S., Chen.Z. and Han.Z.

TITLE Direct Submission

JOURNAL Submitted (30-DEC-1999) Chinese National Human Genome Center at
Shanghai, 351 Guo Shouling Road, Zhangjiang Hi-Tech Park, Pudong,
Shanghai 201203, P. R. China

FEATURES

Location/Qualifiers

source

1..828
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="hypothalamus"
167..544
/codon_start=1
/product="uncharacterized hypothalamus protein HBEX2"
/protein_id="AAF67654.1"
/db_xref="GI:7689028"
/translation="MESKEKRAVNSLSMENAOENKEEOVANKGEPALPLDAGEYC
VPRGNRRFRVQPILQYRWDMHRLGEPQARRENNRIGEEVRLMEKLEKREKOLS
HSLRAVSTDPHHDHDFCLMP"

CDS

BASE COUNT 256 a 170 c 221 g 181 t
ORIGIN

alignment_scores:

Quality: 471.50 Length: 129
Ratio: 4.100 Gaps: 3
Percent Similarity: 89.147 Percent Identity: 70.543

alignment_block:

US-09-327-750D-33 x AF220189

Align seg 1/1 to: AF220189 from: 1 to: 828

```

1 MetGluSerLysAspGlnGlyAlaLysAsn...LeuAsnMetGluAsnAs 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 ATGGAGTCCAAAGAGAACAGCAGTAAACAGTCTCAGCATGGAATAATGC 216
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
16 phisGlnLysLysGluLysGluLysProGlnAspThrIleLysA 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
217 CAACCAAGAAATGAAGAAAGAG.....CAAGTTGCTAATAAAG 257
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
33 rgGluProValValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
258 GGGAGCCCTTGGCCCTCCCT...TTGGATGCTGGTGAATCTGTGCT 304
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyrAr 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
305 AGAGAAATCTAGCGGTTCGGGTTAGGCACCCCATCTCGCAGTATAG 354
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
66 gtrpAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGlu 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
355 ATGGGATATGATGATAGGTGGAGAACACACAGCAAGATGAGAGAAG 404
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
83 LuAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
405 AGAATATGGAAGGATGGGAGGAGGTGAGACAGCTGATGGAAGCTG 454
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
455 AGGGAAGAGCAGTGTAGTCTAGTCTCGGGCAGTCAAGCACTGACCC 504
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116 ohisHisAspHisAspGluPheCysLeuMetPro 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
505 TCACCATGACCATCATGATGAGTTTTCCTTATGCC 541
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: gb_pr:HS198P4

seq_documentation_block:

LOCUS HS198P4 35714 bp DNA PRI 16-DEC-2000
DEFINITION Human DNA sequence from clone RPI-198P4 on chromosome Xq22 Contains
a CpG Island, complete sequence.

ACCESSION AL008708

VERSION AL008708.4 GI:4826445

KEYWORDS HTG; CpG Island.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 35714)

AUTHORS Corby.N.

TITLE
JOURNAL

Direct Submission
Submitted (21-JAN-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk

COMMENT

On May 13, 1999 this sequence version replaced g1:4592117.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.

This sequence was generated from part of bacterial clone contigs of
human chromosome X, constructed by the Sanger Centre Chromosome X
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/ChrX>

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
<http://www.sanger.ac.uk/Projects/C.elegans/wormpep> RPI-198P4 is
from the library RPI-1 constructed at the Roswell Park Cancer
Institute by the group of Pieter de Jong. For further details see
<http://bacpac.med.buffalo.edu/>

VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone

RPI-198P4 it may be shorter because we only sequence overlapping
sections once, or longer because we arrange for a small overlap
between neighbouring submissions.

The true left end of clone LLOXNC01-221F2 is at 35615 in this
sequence. The true right end of clone LLOXNC01-73E8 is at 100 in
this sequence.

FEATURES

source

1. 35714
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q22"
/clone="RPI-198P4"
/clone_lib="RPI-1"

repeat_region
1. 258
/note="L1PA3 repeat: matches 5665. 5925 of consensus"
256. 537
/note="L1PA3 repeat: matches 5391. 5672 of consensus"
1491. 1749
/note="AluSc repeat: matches 39. 297 of consensus"
2182. 3810
/note="L1PB3 repeat: matches 4533. 6150 of consensus"
5169. 6687
/note="L1MA2 repeat: matches 4798. 6308 of consensus"
6976. 7665
/note="L1MA2 repeat: matches 4113. 4798 of consensus"
8039. 9213
/note="L1MA2 repeat: matches 2945. 4113 of consensus"
9523. 10437
/note="L1MA2 repeat: matches 2016. 2945 of consensus"
10736. 11494
/note="L1MA2 repeat: matches 1262. 2016 of consensus"
11807. 12251
/note="L1MCA repeat: matches 310. 755 of consensus"
12259. 12331
/note="HERVH21 repeat: matches 5. 77 of consensus"
12332. 12796
/note="LTR21A repeat: matches 1. 490 of consensus"
16162. 16371
/note="105 copies 2 mer tt 55% conserved"
17121. 17424
/note="MER33 repeat: matches 1. 322 of consensus"

repeat_region
18259. 18290
/note="16 copies 2 mer ca 87% conserved"
18505. 18672
/note="MER31A repeat: matches 1. 181 of consensus"
19478. 19814
/note="MER2 repeat: matches 1. 345 of consensus"
21061. 21101
/note="HERVH21 repeat: matches 42. 85 of consensus"
21247. 21781
/note="MLTIE repeat: matches 44. 568 of consensus"
24465. 24492
/note="Single clone region
Single clone region"
24531. 24562
/note="Weak data
Weak data"
24563
/note="Single clone region
Single clone region"
24711. 24865
/note="Region covered by subclones from a PAC DNA PCR
product only at 12x coverage."
Region covered by subclones from a PAC DNA PCR product only
at 12x coverage."
24758. 25460
/note="CpG island"
/evidence="not_experimental"
24875. 24940
/note="33 copies 2 mer cc 66% conserved"
26795. 26926
/note="66 copies 2 mer ct 63% conserved"
27088. 27133
/note="23 copies 2 mer tg 100% conserved"
27461. 27777
/note="L1MB8 repeat: matches 5851. 6171 of consensus"
27802. 28232
/note="L1MC/D repeat: matches 5192. 5403 of consensus"
28536. 28930
/note="L1MC/D repeat: matches 4802. 5192 of consensus"
29240. 29341
/note="TIGGER1 repeat: matches 484. 587 of consensus"
29603. 30740
/note="TIGGER1 repeat: matches 587. 1757 of consensus"
31027. 31365
/note="TIGGER1 repeat: matches 1757. 2105 of consensus"
31664. 31691
/note="TIGGER1 repeat: matches 2105. 2132 of consensus"
31809. 32015
/note="TIGGER1 repeat: matches 2132. 2246 of consensus"
32323. 32389
/note="TIGGER1 repeat: matches 2246. 2313 of consensus"
34353. 34636
/note="MLT1-INTERNAL repeat: matches 607. 889 of
consensus"
35149. 35376
/note="MLT1-INTERNAL repeat: matches 1001. 1233 of
consensus"
35523. 35670
/note="MLTIC repeat: matches 8. 159 of consensus"
BASE COUNT 9708 a 7094 c 6805 g 12107 t
ORIGIN

alignment_scores:

Quality: 471.50 Length: 129
Ratio: 4.100 Gaps: 3
Percent Similarity: 89.147 Percent Identity: 70.543

alignment_block:

US-09-327-750D-33 x HS198P4/rev

Align seg 1/1 to reverse of: HS198P4 from: 1 to: 35714

```

1 MetGluSerLysAspGlnGlyAlaLysAsn...LeuAsnMetGluAsnAs 16
24358 ATGGAGTCCCAAGAGAACAGCAGTAAACAGTCTCAGCATGGAATGCC 24309

16 pHisGlnLysLysGluGlnLysGluLysProGlnAspThrIleLysA 33
...:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24308 CAACCAAGAAATGAAGAAAGAG...CAAGTTGCTAATAAAG 24268

33 rgGluProValAlaIalProThrPheGluAlaGlyLysAsnCysAlaPro 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24267 GGGAGCCCTTGGCCCTCCCT...TTGGATGCTGGTGAATACTGTGCTG 24221

50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyAr 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24220 AGAGAAATCGTAGGGCGTTCCCGTTAGGCAGCCCATCTCAGTATAG 24171

66 gTrpAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluG 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24170 ATGGATATGATGATAGGCTTGGAGAACACACAGCAAGCATGAGAGAG 24121

83 luAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24120 AGAATATGGAAGGATTTGGGGAGGAGTGGAGACAGCTGATGGAAGCTG 24071

100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24070 AGGAAAGAGGTTGATGATCATGCTGCGGGCAGTCAGACCTGATGCC 24021

116 oHisHisAspHisAspGluPheCysLeuMetPro 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24020 TCACCATGACCATCATGATGAGTTTGGCCTTATGCC 23984

seq_name: gb_pr:AF251053

seq_documentation_block: 744 bp mRNA PRI 15-APR-2001
LOCUS AF251053
DEFINITION Homo sapiens x-linked protein mRNA, complete cds.
ACCESSION AF251053
VERSION AF251053.1 GI:13625167
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Mao,Y., Xie,Y., Zhou,Z., Zhao,W., Zhao,S., Wang,W., Huang,Y.,
Wang,S., Tang,R., Chen,X. and Wu.C.
TITLE Direct Submission
JOURNAL Submitted (29-MAR-2000) Institute of Genetics, School of Life
Sciences, Fudan University, 220 Handan Road, Shanghai 200433, P.R.
China
FEATURES
source Location/Qualifiers
1..744
/organism="Homo sapiens"
/db_xref="taxon:9606"
101..487
/codon_start=1
/product="x-linked protein"
/protein_id="AAK34943.1"
/db_xref="gi:13625168"
/translation="MESKEERALNNLIVENVNDKEQVANKGEPLALPLNVS
EYCVPRGNRRFRVQPIQYRMDIMHRLGEPQARMRENMERIGEEVRLMKLREK
QLSHSLAVSTDPPPHDHIDEFLMP"
BASE COUNT 222 a 134 c 208 g 180 t
ORIGIN

alignment_scores:
Quality: 468.00 Length: 129
Ratio: 4.000 Gaps: 2
Percent Similarity: 90.698 Percent Identity: 68.217

alignment_block:
US-09-327-750D-33 x AF251053
Align seg 1/1 to: AF251053 from: 1 to: 744

1 MetGluSerLysAspGlnGlyAla...LysAsnLeuAsnMetGluAsnAs 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 ATGGAGTCCCAAGAGAACAGCAGTAAACAGTCTCAGCATGGAATGCT 150

16 pHisGlnLysLysGluGlnLysGluLysProGlnAspThrIleLysA 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 CAACCAAGAAATGAAGAAAGAG...CAAGTTGCTAATAAAG 200

33 rgGluProValAlaIalProThrPheGluAlaGlyLysAsnCysAlaPro 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 GGGAGCCCTTGGCCCTCCCT...TTGAATGTTAGTGAATACTGTGCTG 247

50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyAr 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
248 AGAGAAACCGTAGCGGTTCCCGTTAGGCAGCCCATCTCGCATATAG 297

66 gTrpAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluG 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 ATGGACATATGATGATAGGCTTGGAGAGCCACAGCAAGCATGAGAGAG 347

83 luAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
348 AGAATATGGAAGGATTTGGGGAGGAGTGGAGACAGCTGATGGAAGCTG 397

100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
398 AGGAAAGAGCAGTTGATGATCATGCTGCGGGCAGTCAGACCTGATGCC 447

116 oHisHisAspHisAspGluPheCysLeuMetPro 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
448 TCACCATGACCATCATGATGAGTTTGGCCTTATGCC 484

seq_name: gb_pat:AX078272

seq_documentation_block: 898 bp DNA PAT 22-FEB-2001
LOCUS AX078272
DEFINITION Sequence 76 from Patent WO0107471.
ACCESSION AX078272
VERSION AX078272.1 GI:13157963
KEYWORDS human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Hattman,J.L., Lal,P., Tang,Y.T., Yue,H., Au-Young,J., Bandman,O.,
Azimzadeh,F., Yang,J., Lu,D.A., Baughn,M.R., Patterson,C. and Shah,P.
TITLE Cell cycle and proliferation proteins
JOURNAL Patent: WO 0107471-A 76 01-FEB-2001;
Incyte Genomics, Inc. (US)
FEATURES
source Location/Qualifiers
1..898
/organism="Homo sapiens"
/db_xref="taxon:9606"
/notes="Incyte ID No: 5664154CB1"
BASE COUNT 250 a 186 c 251 g 211 t
ORIGIN

alignment_scores:
Quality: 468.00 Length: 129
Ratio: 4.000 Gaps: 2
Percent Similarity: 90.698 Percent Identity: 68.217

alignment_block:
US-09-327-750D-33 x AX078272
Align seg 1/1 to: AX078272 from: 1 to: 898

```

```

1 MetGluSerLysAspGlnGlyAla...LysAsnLeuAsnMetGluAsnAs 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 ATGGAGTCCAAAGAGGACGAGGTTAAACAATCTCATCGTGGAAAATGT 303
16 PhISGLnLysLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
304 CACCCAGAAAGATGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAG 353
33 rGluPValValAlaProThrPheGluAlaGlnGlnGlnGlnGlnGlnGln 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
354 GGGAGCCCTGGCCCTACCT...TTGAATGTTAGTGAATCTGTGCGCT 400
50 ArgGlyGlyArgArgArgPheArgValArgGlnProIleSerHisTyrAr 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
401 AGAGAAACCGTAGGCGGTTCCCGTTAGGCAGCCCATCCTGCAGATAG 450
66 gTriPAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGlu 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
451 ATGGACATAATGCATAGCTTGGAGAGCCACAGGCAAGGATGAGAGAGG 500
83 LuAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGlnLysLeu 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
501 AGAATATGGAAGAGATGGGAGGAGGTGAGACAGCTGATGGAAGCTG 550
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
551 AGGAAAAGCAGTTGAGTCATAGCTCTCGCGGCGAGTCAGCACTGATCCC 600
116 OHISHisAspHisHisAspGluPheCysLeuMetPro 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
601 TCACCATGACCATCAGCATGAGTTTGGCTTATGCC 637

```

seq_name: gb_pr:HSV870H8

```

seq_documentation_block:
LOCUS      HSV870H8      31321 bp      DNA      PRI      23-NOV-1999
DEFINITION Human DNA sequence from cosmid V870H8, between markers DXS366 and
            DXS87 on chromosome X contains ESTs.
ACCESSION  270233
VERSION    270233.1
KEYWORDS   X.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 31321)
AUTHORS    Whiteley, M.
TITLE      Direct Submission
JOURNAL    CB10 1R0, UK. E-mail enquires: humquery@sanger.ac.uk
COMMENT    IMPORTANT: This sequence is not the entire insert of clone V870H8.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone V870H8 is at 1 in this sequence. The
            true left end of clone V693A8 is at 31221.
            V870H8 is from the human chromosome X-specific cosmid library.
FEATURES   Location/Qualifiers
            1..31321
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="X"
             /map="X"
             /clone="GHC-870H8"
             /clone_lib="SCcv"
            repeat_region 475..701
             /note="L1 element fragment"
            repeat_region 802..988
             /note="L1 element fragment"
            misc_feature 2533..4138
             /note="match: multiple ESTs"
            repeat_region 4189..4246
             /note="29 copies of 2 mer 91 & conserved"

```

```

repeat_region 7496..7874
 /note="L1 element fragment"
repeat_region 8022..8108
 /note="L1 element fragment"
repeat_region 8302..8493
 /note="L1 element fragment"
repeat_region 8585..8671
 /note="L1 element fragment"
repeat_region 8769..9010
 /note="L1 element fragment"
repeat_region 9259..9384
 /note="L1 element fragment"
repeat_region 9624..9731
 /note="MLTIC element fragment"
repeat_region 9781..9948
 /note="MLTIC element fragment"
repeat_region 9890..9948
 /note="MLTIB element fragment"
repeat_region 10418..10592
 /partial
 /note="Alu repeat: matches 298..108 of consensus"
repeat_region 10593..10688
 /partial
 /note="Alu repeat: matches 96..1 of consensus"
repeat_region 10698..10781
 /note="L1 element fragment"
repeat_region 11112..11403
 /note="Alu repeat: matches 1..308 of consensus"
repeat_region 11922..12004
 /note="L1 element fragment"
repeat_region 12062..12308
 /note="L1 element fragment"
repeat_region 12423..12707
 /partial
 /note="Alu repeat: matches 302..1 of consensus"
repeat_region 12710..12916
 /note="L1 element fragment"
repeat_region 12979..13081
 /note="L1 element fragment"
repeat_region 13473..13532
 /note="MLT2A1 element fragment"
repeat_region 15871..15928
 /note="L1 element fragment"
repeat_region 18180..18417
 /note="L1 element fragment"
repeat_region 18778..19056
 /note="Alu repeat: matches 1..308 of consensus"
repeat_region 22279..22350
 /note="L1 element fragment"
repeat_region 22449..22995
 /note="L1 element fragment"
repeat_region 23239..23374
 /note="L1 element fragment"
repeat_region 23371..23606
 /note="MLTIB element fragment"
repeat_region 23395..23541
 /note="MLTIA element fragment"
repeat_region 23705..23778
 /note="MLTIA element fragment"
repeat_region 23805..23960
 /note="L1 element fragment"
repeat_region 24011..24541
 /note="L1 element fragment"
repeat_region 24638..24770
 /partial
 /note="Alu repeat: matches 150..1 of consensus"
repeat_region 24829..30975
 /note="L1 element fragment"
BASE COUNT 9462 a 6147 c 6129 g 9583 t
ORIGIN
alignment_scores:

```



```

repeat_region /note="L2 repeat: matches 1599. .2175 of consensus"
18706. .18755
repeat_region /note="L1M3 repeat: matches 6034. .6083 of consensus"
18756. .19236
repeat_region /note="MER66B repeat: matches 1. .486 of consensus"
19337. .21108
/note="MER66-internal repeat: matches 4808. .6676 of
consensus"
21109. .21163
repeat_region /note="L1M2B repeat: matches 391. .445 of consensus"
21164. .21555
repeat_region /note="MER66B repeat: matches 3. .391 of consensus"
21557. .21662
repeat_region /note="L1M3B repeat: matches 6080. .6185 of consensus"
21665. .21761
repeat_region /note="L1M repeat: matches 5299. .5392 of consensus"
21762. .21929
repeat_region /note="L1M8B repeat: matches 5902. .6063 of consensus"
21930. .22212
repeat_region /note="AluSg repeat: matches 3. .284 of consensus"
22213. .22321
repeat_region /note="L1M8B repeat: matches 6063. .6173 of consensus"
22333. .22740
repeat_region /note="L1MD2 repeat: matches 5867. .6288 of consensus"
22744. .23094
repeat_region /note="L1MD repeat: matches 2. .342 of consensus"
23218. .23524
repeat_region /note="AluSg repeat: matches 1. .305 of consensus"
23603. .23947
repeat_region /note="L1M4 repeat: matches 4916. .5184 of consensus"
23948. .24218
repeat_region /note="AluJo repeat: matches 1. .282 of consensus"
24219. .24510
repeat_region /note="L1M4 repeat: matches 4601. .4916 of consensus"
24508. .24621
repeat_region /note="L1M4 repeat: matches 4374. .4488 of consensus"
24650. .25077
repeat_region /note="MLTIC repeat: matches 1. .466 of consensus"
25078. .25914
repeat_region /note="L1M4 repeat: matches 3497. .4346 of consensus"
25920. .26741
repeat_region /note="L1M4 repeat: matches 2202. .3065 of consensus"
26762. .26906
repeat_region /note="L1P7 repeat: matches 6001. .6145 of consensus"
26907. .27135
repeat_region /note="L1P5 repeat: matches 5674. .5902 of consensus"
27266. .27509
repeat_region /note="L1M4 repeat: matches 2164. .2080 of consensus"
27557. .27769
repeat_region /note="L1M3C repeat: matches 253. .460 of consensus"
28120. .28167
repeat_region /note="L2 repeat: matches 2703. .2750 of consensus"
29043. .29349
repeat_region /note="MER66B repeat: matches 1. .341 of consensus"
30389. .30452
/note="16 copies 4 mer tata.89% conserved"
complement(30502)
/note="dJ79p11.1"
/note="dJ79p11.1"
complement(join(30502. .31137.31514. .31589.32016. .>32134))
/note="match: ESTs: Em:AA317587 Em:AA358632 Em:A1214048
Em:A1929470 Em:AA319216 Em:AA317625 Em:D81392 Em:AA317418
Em:AA365528 Em:C15186 Em:A1291270 Em:AA160660 Em:AA612605
Em:AW160832 Em:AA384405 Em:AA160657 Em:AA081744
Em:A1291126 Em:AA35591 Em:AA384268 Em:A1929703 Em:A1302026
Em:W19547 Em:A1141727 Em:AA946933 Em:A1141583 Em:H61106
Em:W40581 Em:A1826617 Em:AA706545 Em:AA464771 Em:AA862327
Em:W48380 Em:AA641652 Em:AA384269 Em:AA527696 Em:W48778
Em:AA633989 Em:A1659486 Em:AW167301 Em:A1929428
Em:AW379411 Em:C15948 Em:A1745669 Em:AW024587 Em:AW379207;
match: proteins: Tr:O9Y516 Tr:O9Y517"
/evidence="not_experimental"
/product="dJ79p11.1 (novel protein similar to mouse Bex2

```

```

(brain-expressed X-linked protein 2))"
complement(30502. .32134)
/note="dJ79p11.1"
30507. .30988
/note="match: STS: Em:G24641; match: STS: Em:T86927"
30508. .30852
/note="match: STS: Em:R43117; match: STS: Em:G23964"
complement(30518. .30523)
/note="dJ79p11.1"
30541. .30922
/note="match: STS: Em:H61107"
31543. .32262
/note="CpG island"
/evidence="not_experimental"
31647. .31730
/note="42 copies 2 mer cc 64% conserved"
complement(32655. .33316)
/note="match: GSS: Em:AQ489303"
33027. .34514
/note="L1M2 repeat: matches 4631. .6155 of consensus"
34528. .34786
/note="L1M3A repeat: matches 4404. .4664 of consensus"
34787. .35085
/note="AluSg repeat: matches 1. .300 of consensus"
35086. .35412
alignment_scores:
Quality: 468.00 Length: 129
Ratio: 4.000 Gaps: 2
Percent Similarity: 90.698 Percent Identity: 68.217
alignment_block:
US-09-327-750D-33 x AL133348/rev
Align seg 1/1 to reverse of: AL133348 from: 1 to: 40584
1 MetGluSerLysAspGlnGlyAla...LysAsnLeuAsnMetGluAsnAs 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31132 ATGGAGTCCAAAGAGGACGCGCTTAACAATCTCATCTGGAATGT 31083
16 phiSGlnLysLysGluLysGluLysGluLysProGlnAspThrIleLysA 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
31082 CAACAGGAAATGATGAAAGATGAAAGGACCAAGTTCCTAATAAG 31033
33 rGluProValValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
31032 GGGAGCCCTTGGCCCTACTT...TTGAATGTTAGTAATCTGTGCT 30986
50 ATGGTGLVATGATGATGATGATGATGATGATGATGATGATGATGATGAT 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
30985 AGAGGAAACCTAGCGGCTCCGCTTAGCGAGCCATCTCGCAGTATAG 30936
66 gtrAspLeuMetHisArgValGlyGluProGlnGlnArgMetArgGluG 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
30935 ATGGGACATATATGATAGCTTGGAGAGCCACAGGCAAGGATGAGAGG 30886
83 luAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
30885 AGAATATGAAAGGATGGGGAGGAGGTGACACAGCTGATGAAAGCTG 30836
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
30835 AGGGAAGACAGTTCAGTCACTAGTTTCGGGCGCAGTATGATCCCC 30786
116 ohHisAspHisHisAspGluPheCysLeuMetPro 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
30785 TCACCATGACCATCAGTATGAGTTTGCCTTATCCCC 30749
seq_name: gb_hgt:HS080B1
seq_documentation_block:
LOCUS HS080B1 41029 bp DNA HTG
DEFINITION Homo sapiens chromosome X clone LL0XNC01-80B1 map q22.1-22.3, ***

```


Align seg 1/1 to: AF237783 from: 1 to: 791

```

1 MetGluSerLysAspGlnGlyAlaLysAsp...LeuAsnMetGluAsnAs 16
|||||
172 ATGGAGTCCAAAGAACTAGCAGTAACAGTCTCAGCATGGAATAATGC 221
16 phiGlnLysLysGluGluLysGluGluLysProGlnAspThrIleLysA 33
:|||||
222 CAACCAAGAAATGAAGAAAGAG.....CAAGTTGCTAATAAAG 262
33 rgGluProValAlaProThrPheGluAlaClyLysAsnCysAlaPro 49
|||||
263 GGGAGCCCTGGCCCTCCCT...TTGATGCTGGTGAATACGTGTGCT 309
50 ArgGlyArgArgPheArgValArgGlnProIleSerHisTyrAr 66
|||||
310 AGAGAAATCGTAGCGGTTCCGCGTTAGCAGCCCATCTCGCAGTATAG 359
66 gTIPAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluG 83
|||||
360 ATGGATATGATGCTATAGGCTTGGAGAACACAGGCAAGGTAGAGAG 409
83 luAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99
|||||
410 AGAATATGGAAGGATTTGGGAGGGGTGAGACAGCTGATGGAAGCTG 459
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
|||||
460 AGGAAAGAGTGTAGTATAGTCTGCGGGCAGTCAGCAGTACGCCCC 509
116 oHISHisAspHisAspGluPheCysLeuMetPro 128
|||||
510 TCACCATGACCATCATGATGAGTTTGCCTTATGCC 546

```

seq_name: gb_sts:G24641

```

seq_documentation_block: 504 bp DNA STS 31-MAY-1996
LOCUS G24641
DEFINITION human STS WI-11354, sequence tagged site.
ACCESSION G24641
VERSION G24641.1 GI:1344967
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human STS derived from sequences in dbEST and the Unigene
collection.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Hudson.T.
Whitehead Institute/MIT Center for Genome Research; Physically
Mapped STS
Unpublished (1995)
JOURNAL
COMMENT

```

Contact: Thomas Hudson
 Whitehead Institute/MIT Center for Genome Research
 Whitehead Institute for Biomedical Research
 9 Cambridge Center, Cambridge MA 02142 USA
 Tel: 617 252 1900
 Fax: 617 252 1902
 Email: thudson@genome.wi.mit.edu

Primer A: TTTTGTGCTTTTAAAGTGC
 Primer B: TTACGTGGGTCTCCTATTACCA
 STS size: 150
 PCR Profile:

Presoak:
 Denaturation:
 Annealing: 56 degrees C
 Polymerization:
 PCR Cycles: 35
 Thermal Cycler:
 Protocol:
 Template: 10 ng

Primer: each 5 pM
 dNTPs: each 4 nM
 Taq Polymerase: 0.025 units/ul
 Total Vol: 20 ul

Buffer:

MgCl2: 1.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 9.3

Derived from dbEST (genbank accession #86927).

```

FEATURES
    source            1..504
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
    STS
    primer_bind       1..150
    primer_bind       1..22
    primer_bind       complement(129..150)
    BASE COUNT       134 a 118 c 101 g 144 t 7 others
    ORIGIN

```

alignment_scores:
 Quality: 311.00 Length: 85
 Ratio: 4.260 Gaps: 3
 Percent Similarity: 85.882 Percent Identity: 72.941

alignment_block:

US-09-327-750D-33 x G24641/rev ..

Align seg 1/1 to reverse of: G24641 from: 1 to: 504

```

47 CysAlaProArgGly.GlyArgArgArgPhe.ArgValArgGlnProIle 62
|||
503 TGTCCTCCCTANAGGAAACCGTAGCGGTTCCCTCCCTTAGGCANCCAC 454
63 SerHisTyrArgTIPAspLeuMetHisArgValGlyGlu.ProGlnGlyA 79
:::
453 CCGCAGTANAGATGGGATATGATGATGATGATGATGATGATGATGATG 404
79 rgMetArgGluGluAsnValGlnArgPheGlyGluAspMetArgGlnLeu 95
|||||
403 GGATGAGAGAGAGATATGGAAGGATTTGGGAGGAGGTGAGACAGCTG 354
96 MetGluLysLeuArgGluArgGlnLeuSerHisSerLeuArgAlaValSe 112
|||||
353 ATGAAAGAGCTGAGGAAAGAGAGTTCAGTCANAGTCTCGGGCAGTCAG 304
112 rThrAspProHisHisAspHisHisAspGluPheCysLeuMetPro 128
|||||
303 CACTGACCCCTCACCATGACCATCATGATGAGTTTCTGCTTATGCC 255

```



```
;
; TOPOLOGY: both
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 326..6277
; US-08-836-325-9

alignment_scores:
  Quality: 72.50 Length: 58
  Ratio: 1.812 Gaps: 4
  Percent Similarity: 68.966 Percent Identity: 36.207

alignment_block:
  US-09-327-750D-34 x US-08-836-325-9/rev ..
  Align seg 1/1 to reverse of: US-08-836-325-9 from: 1 to: 6452

25 LysAlaSerLysGlnSerGluGluGlu.....SerHisHisLeuG1 38
|||||:|||||:|||||:|||||:|||||:
3220 AAGAGCCAGAAACAGGTTCCAGCACCAAGGTTCCCAATCACCATGACCA 3171

38 uGluValGluAsnLysLysProGlyGlyAsnValArgGlyValArgA 55
|||||:|||||:|||||:|||||:|||||:
3170 TCATGTAAACATTAAGGCACATGTT.....CTGCCCGCGACCTC...C 3130

55 rgLeuValProAsnProAlaLeuProAsnArgHisValAspHis 71
|||||:|||||:|||||:|||||:|||||:
3129 ATGCAGTCCACATGCTCTATCCACTCCCCACACAGCAGCTCGGAACAC 3080

72 SerGluGlyGly...GluGluVal 78
|||||:|||||:|||||:|||||:
3079 GATGAGGAGGAGGTGGGAAGATC 3056

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-892-770-1

seq_documentation_block:
; Sequence 1, Application US/08892770
; Patent No. 5962670
; GENERAL INFORMATION:
; APPLICANT: Walling, Linda L.
; APPLICANT: Pautot, Veronique
; APPLICANT: Gu, Yong-Qiang
; APPLICANT: Chao, Wun Shaw
; TITLE OF INVENTION: Improved Promoters for Enhancing Plant
; TITLE OF INVENTION: Productivity
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/892,770
; FILING DATE: 15-JUL-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Quine, Jonathan A.
; REGISTRATION NUMBER: P-41,261
; REFERENCE/DOCKET NUMBER: 023070-072100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 935 base pairs

;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY:
; LOCATION: 1..935
; OTHER INFORMATION: /note= "tomato acidic leucine
; OTHER INFORMATION: aminopeptidase 1 (Lapal) promoter
; OTHER INFORMATION: region"
; US-08-892-770-1

alignment_scores:
  Quality: 72.00 Length: 60
  Ratio: 2.250 Gaps: 2
  Percent Similarity: 53.333 Percent Identity: 28.333

alignment_block:
  US-09-327-750D-34 x US-08-892-770-1 ..
  Align seg 1/1 to: US-08-892-770-1 from: 1 to: 935

62 TrpAlaIleProAsnArgHisValAspHisSerGluGlyGlyGluGluVa 78
|||||:|||||:|||||:|||||:|||||:
438 TGGTGTGTGTGAATGAGATTATCGATCATATAAGAGAGGGGAGAGAT 487

78 lGlyArgPheValGlyGlnValMet.GluAlaLysArgHis..... 91
|||||:|||||:|||||:|||||:
488 GAACAGTTTTTTAAATAAATTTTGGCCCCAGAAAAAACATTCCACGCGT 537

92 .....Serly 93
|||||

538 GACTTTAGTCATTACATCACCTTGCCTGACAGAGAAATCGTCAAA 587
|||||
93 sGluGlnGlnMetArgProTyrThrarg 102
|||||
588 AAATGACACATGAGACCTTACATAAGG 615

seq_name: /cgn2_5/ptodata/2/ina/6A_COMB.seq:US-08-726-214-5

seq_documentation_block:
; Sequence 5, Application US/08726214
; Patent No. 6107076
; GENERAL INFORMATION:
; APPLICANT: Tang, Wei-Jen
; APPLICANT: Gliman, Alfred G.
; TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYL CYCLASE
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,214
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/005,498
; FILING DATE: 04-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: UTSD:450
```


Allan seq 1/1 to: US-08-841-483-3 from: 1 to: 3490


```

; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/807.043B
; FILING DATE: 19911212
; CLASSIFICATION: 434
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764.364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728.838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705.702
; FILING DATE: 23-May-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5342774man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 253.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4698 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: singular
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-07-807-043B-5

alignment_scores:
  Quality: 71.00 Length: 93
  Ratio: 1.365 Gaps: 4
  Percent Similarity: 55.914 Percent Identity: 27.957

alignment_block:
US-09-327-750D-34 x US-07-807-043B-5/rev ..
Align seg 1/1 to reverse of: US-07-807-043B-5 from: 1 to: 4698
16 LysAspLysLysAsnLysGlyLysAlaSer.LysGlnSerGluG 32
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2134 AGGGACAGGAAGGCAAGAGGGGGGCGCCAGCGAGGGGAGGAC 2085
32 luGluSerHisHisLeuGluGluValGluAsnLys.....Lys 44
::||| ||| :: |||:|||||:
2084 AGGAAAGGCACACAGAGGGAAGCGGAGGGGGGGGATCTTGAAAA 2035
45 ProGlyGlyAsnValArgLysValArgLysValProAsnPheLe 61
|||||:|||||:|||||:|||||:|||||:|||||:
2034 CAGGTGCAAAAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAA 1990
61 uTrpAlaIleProAsnArgHisValAspHisSerGluGlyGluGlu 78
|||||:|||||:|||||:|||||:|||||:|||||:
1989 .....CCAAACAAACAAACAAACAAACAAACAAACAAAC 1950
78 alGly.....ArgPheValGlyGlnValMetGluAlaLys 89
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1949 CAGGAGGAGGCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGT 1900
90 ArgHisSerLysGluGlnMetArg 98
::: |||:|||||:
1899 AAGGCTGCTAAAGTCGACGAAAGAGG 1873

seq_name: /cnp2.6/ptodata/2/ina/5A_COMB.seq:US-08-299-849B-5
seq_documentation_block:
; Sequence 5, Application US/08299849B
; Patent No. 5612201
; GENERAL INFORMATION:
; APPLICANT: De Plaen, Etienne; Boon-Falleur, Thierry;
; APPLICANT: Leth, Bernard; Szikora, Jean-Pierre; De Smet, Charles;
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful In
; APPLICATION OF INVENTION: Determining Expression Of A Tumor Antigen Precursor
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felte & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,849B
; FILING DATE: 1-SEPTEMBER-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/037,230
; FILING DATE: 26-MARCH-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 9-JULY-1991
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-May-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5612201man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5355
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4698 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-299-849B-5

alignment_scores:
  Quality: 71.00 Length: 93
  Ratio: 1.365 Gaps: 4
  Percent Similarity: 55.914 Percent Identity: 27.957

alignment_block:
US-09-327-750D-34 x US-08-299-849B-5/rev ..
Align seg 1/1 to reverse of: US-08-299-849B-5 from: 1 to: 4698
16 LysAspLysLysAsnLysGlyLysAlaSer.LysGlnSerGluG 32
:::|||||:|||||:|||||:|||||:|||||:|||||:
2134 AGGGACAGGAAGGCAAGAGGGGGGCGCCAGCGAGGGGAGGAC 2085
32 luGluSerHisHisLeuGluGluValGluAsnLys.....Lys 44
::||| ||| :: |||:|||||:
2084 AGGAAAGGCACACAGAGGGAAGCGGAGGGGGGGGATCTTGAAAA 2035
45 ProGlyGlyAsnValArgLysValArgLysValProAsnPheLe 61
|||||:|||||:|||||:|||||:|||||:|||||:
2034 CAGGTGCAAAAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAA 1990
61 uTrpAlaIleProAsnArgHisValAspHisSerGluGlyGluGlu 78
|||||:|||||:|||||:|||||:|||||:|||||:
1989 .....CCAAACAAACAAACAAACAAACAAACAAACAAAC 1950
78 alGly.....ArgPheValGlyGlnValMetGluAlaLys 89
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1949 CAGGAGGAGGCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGT 1900
90 ArgHisSerLysGluGlnMetArg 98
::: |||:|||||:
1899 AAGGCTGCTAAAGTCGACGAAAGAGG 1873

seq_name: /cnp2.6/ptodata/2/ina/5A_COMB.seq:US-08-299-849B-5
seq_documentation_block:
; Sequence 5, Application US/08299849B
; Patent No. 5612201
; GENERAL INFORMATION:
```


Tue Mar 12 09:01:44 2002

```
45 ProGlyGlyAsnValArgArgLysValArgArgLeuValProAsnPheLe 61
|||:||||: |||:||||: |||:||||: |||:||||: |||:||||: |||:||||:
2034 CAAGGTGCAAAAAAAAAAAAAAAAAAGCAAAACCAACCAACCAACAA..... 1990

61 uTrpAlaIleProAsnArgHisValAspHisSerGluGlyGluGluV 78
|||:||||: |||:||||: |||:||||: |||:||||: |||:||||: |||:||||:
1989 .....CCAAACAACAAAAAGGTGAGCTGGAGGTGGAGGGGAGC 1950

78 alGly.....ArgPheValGlyGlnValMetGluAlaLys 89
|||:||||: |||:||||: |||:||||: |||:||||: |||:||||: |||:||||:
1949 CAGGAGGAAGGCACAAAAAGGGGACGCGGCAGAAAAAGCAGGAGGAGGT 1900

90 ArgHisSerLysGluGlnGlnMetArg 98
|||:||||: |||:||||: |||:||||: |||:||||: |||:||||: |||:||||:
1899 AAGCTGCTAAAAGTCGAGCAAGAGG 1873
```

/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.AAH54129 +	76.00	133.47	23.35	354.
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.AAH93013 -	75.50	149.04	3.17	700
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.AAC73887 -	75.50	142.27	7.55	135.
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.AAA72409 +	75.00	142.42	7.41	119.
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.AAS00010 -	75.00	133.35	23.71	289.

```
seq_name: 791b2a279c94data/geneseq/geneseqn/AAH03517
seq_documentation_block:
ID   AAH03517 standard; cDNA; 865 BP.
XX
XX
AC   AAH03517;
XX
XX   26-JUN-2001 (first entry)
XX
XX   Human cDNA clone (5'-primer) SEQ ID NO:352.
XX
DE
```

XX	EP1074617-A2.
PN	
XX	
XX	07-FEB-2001.
PD	
XX	
XX	28-JUL-2000; 2000EP-0116126.
PF	
XX	
XX	29-JUL-1999; 99JP-0248036.
PR	27-AUG-1999; 99JP-0300253.
PR	11-JAN-2000; 2000JP-0118776.
PR	02-MAY-2000; 2000JP-0183767.
PR	09-JUN-2000; 2000JP-0241899.
PR	

Sequence	Strd Orig	zScore	EScore Len	Documentation
/SID2/cgcdata/geneseq/NA2001.DAT:AAH03517	+ 372.00	800.02	1.7e-36	865
/SID2/cgcdata/geneseq/NA2001.DAT:AAH13750	+ 372.00	796.41	2.8e-36	1229
/SID2/cgcdata/geneseq/NA2001.DAT:AAH83548	+ 372.00	795.34	3.2e-36	1364
/SID2/cgcdata/geneseq/NA2001.DAT:AAF23529	+ 171.50	358.04	7.2e-12	891
/SID2/cgcdata/geneseq/NA2000.DAT:AAF21748	+ 170.00	354.44	1.1e-11	917
/SID2/cgcdata/geneseq/NA2001.DAT:AAF23528	+ 144.00	299.94	1.2e-08	700
/SID2/cgcdata/geneseq/NA2000.DAT:AAH01004	+ 143.00	301.51	1.0e-08	485
/SID2/cgcdata/geneseq/NA2000.DAT:AAH01005	+ 136.00	285.14	8.3e-08	532
/SID2/cgcdata/geneseq/NA2001.DAT:AAH75810	+ 111.00	226.31	0.0002	767
/SID2/cgcdata/geneseq/NA2001.DAT:AAH5851	+ 111.00	225.10	0.0002	862
/SID2/cgcdata/geneseq/NA2001.DAT:AAH95611	+ 111.00	224.68	0.0002	898
/SID2/cgcdata/geneseq/NA2001.DAT:AAH45143	+ 109.50	222.67	0.0003	792
/SID2/cgcdata/geneseq/NA2000.DAT:AAH03880	+ 103.00	210.20	0.0012	662
/SID2/cgcdata/geneseq/NA2000.DAT:AAH03879	+ 103.00	209.65	0.0013	698
/SID2/cgcdata/geneseq/NA2001.DAT:AAI60367	+ 99.00	198.72	0.0054	858
/SID2/cgcdata/geneseq/NA2000.DAT:AAH01089	+ 98.50	199.83	0.0047	692
/SID2/cgcdata/geneseq/NA2000.DAT:AAH01003	+ 97.50	203.23	0.0030	401
/SID2/cgcdata/geneseq/NA1999.DAT:AAH23519	+ 92.50	143.81	6.20	44453
/SID2/cgcdata/geneseq/NA2001.DAT:AAI12021	+ 89.50	189.45	0.0178	276
/SID2/cgcdata/geneseq/NA2001.DAT:AAI13401	+ 89.50	189.45	0.0178	276
/SID2/cgcdata/geneseq/NA2001.DAT:AAI05905	+ 89.50	189.45	0.0178	276
/SID2/cgcdata/geneseq/NA2001.DAT:AAI13308	+ 87.50	180.85	0.0536	415
/SID2/cgcdata/geneseq/NA2001.DAT:AAI34561	+ 87.50	180.85	0.0536	415
/SID2/cgcdata/geneseq/NA2001.DAT:AAI03209	+ 87.50	180.85	0.0536	415
/SID2/cgcdata/geneseq/NA2000.DAT:AAZ57839	+ 83.50	158.39	0.9561	1566
/SID2/cgcdata/geneseq/NA2000.DAT:AAH10225	+ 83.50	114.98	250.27	10674
/SID2/cgcdata/geneseq/NA2000.DAT:AAH06225	+ 81.00	150.02	2.80	2069
/SID2/cgcdata/geneseq/NA1999.DAT:AAH80625	+ 80.00	159.18	0.8631	762
/SID2/cgcdata/geneseq/NA2001.DAT:AAH28355	+ 80.00	141.62	262.02	52216
/SID2/cgcdata/geneseq/NA2000.DAT:AAZ46153	+ 78.00	137.19	14.49	3786
/SID2/cgcdata/geneseq/NA2000.DAT:AAH77252	+ 78.00	136.98	14.89	3866
/SID2/cgcdata/geneseq/NA2001.DAT:AAH08642	+ 78.00	136.82	15.19	3925
/SID2/cgcdata/geneseq/NA1995.DAT:AAH086237	+ 78.00	135.42	18.19	4500
/SID2/cgcdata/geneseq/NA1995.DAT:AAH86236	+ 78.00	135.31	18.45	4548
/SID2/cgcdata/geneseq/NA2000.DAT:AAH10222	+ 77.50	131.37	30.58	5994
/SID2/cgcdata/geneseq/NA2000.DAT:AAH39709	+ 77.00	150.83	2.52	811
/SID2/cgcdata/geneseq/NA2001.DAT:AAH22774	+ 77.00	147.68	3.77	1102
/SID2/cgcdata/geneseq/NA2001.DAT:AAI12149	+ 76.00	156.05	1.29	394
/SID2/cgcdata/geneseq/NA2001.DAT:AAI46784	+ 76.00	156.05	1.29	394
/SID2/cgcdata/geneseq/NA2001.DAT:AAI07190	+ 76.00	156.05	1.29	394

05 JUN 2000, 2000JF 04418593.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
Ishii S, Sujiyama T, Wakamatsu A, Nagai K, Otsuki T;
WPI: 2001-318749/34.

primer sets for synthesizing polynucleotides, particularly the 5602
full-length cDNAs defined in the specification, and for the detection
and/or diagnosis of the abnormality of the proteins encoded by the
full-length cDNAs -

Claim 1: SEQ ID 352; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602
full-length cDNAs defined in the specification. Where a primer set
comprises: (a) an oligo-dT primer and an oligonucleotide complementary
to the complementary strand of a polynucleotide which comprises one of
the 5602 nucleotide sequences defined in the specification, where the
oligonucleotide comprises at least 15 nucleotides; or (b) a combination
of an oligonucleotide comprising a sequence complementary to the
complementary strand of a polynucleotide which comprises a 5'-end
sequence and an oligonucleotide comprising a sequence complementary to a
polynucleotide which comprises a 3'-end sequence, where the
oligonucleotide comprises at least 15 nucleotides and the combination of
the 5'-end sequence/3'-end sequence is selected from those defined in
the specification. The primer sets can be used in antisense therapy and
in gene therapy. The primers are useful for synthesizing polynucleotides,
particularly full-length cDNAs. The primers are also useful for the
detection and/or diagnosis of the abnormality of the proteins encoded by
the full-length cDNAs. The primers allow obtaining of the full-length
cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
AAH13633 to AAH18742 represent human cDNA sequences; AA92446 to
AA95893 represent human amino acid sequences; and AAH13629 to AAH13632
represent oligonucleotides, all of which are used in the exemplification
of the present invention.

Sequence 865 BP: 239 A: 177 C: 224 G: 221 T: 4 other:

alignment_scores:
 Quality: 372.00 Length: 120
 Ratio: 3.758 Gaps: 2
 Percent Similarity: 82.500 Percent Identity: 60.000

alignment_block:
 US-09-327-750D-34 x AAH03517 ..

Align seg 1/1 to: AAH03517 from: 1 to: 865

```

1 MetAlaSerLysValLysGlnValIleLeuAspLeuThrValGluLysAs 17
||||| ||||| : : : : : ||||| : : : : :
191 ATGGAGTCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240

17 pLysLysAsnLysLysGlyGly...LysAlaSerLysGlnSerGluGlu 33
: : : : : : : : : : : : : : : : : : : : : : : : : : :
241 CCAACAAGAAACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 290

33 LuSerHisHisLeuGluValGluValGluValGluValGluValGlu 49
||||| ||||| : : : : : ||||| : : : : :
291 AATCCCGCCATTGGGAGGGGGTGAAGCCAGAGGAGGAGGAGGAGGAGG 340

50 ArgArg...LysValArgArgLeuValProAsnPheLeuTrpAlaIlePr 65
||||| ||||| : : : : : ||||| : : : : :
341 AGCGGGGGCGAGTATGAGCGAGTATGAGCGAGTATGAGCGAGTATGAG 390

65 OAsnArgHisValAspHisSerGluGlyGlyGluValGluValGluVal 82
||||| ||||| : : : : : ||||| : : : : :
391 TAATAGGCATATTGAGCAGACATGAGCGAGGAGGAGGAGGAGGAGGAG 440

82 aLcGlyGlnValMetGluAlaLysArgHisSerLysGluGlnGlnMetArg 98
||||| ||||| : : : : : ||||| : : : : :
441 TAGGCAGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 490

99 ProTyrThrArgPheArgThrProGluProAspAsnHisTyrAspPheCy 115
||||| ||||| : : : : : ||||| : : : : :
491 CACTATATGCGCTTCCAACTCTGTGACCTGACCACTTATGACTTTTG 540

115 sleullePro 118
||||| |||||
541 CCTCATACCT 550

```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH13750

seq_documentation_block:
 ID: AAH13750 standard; CDNA: 1229 BP.

XX AC AAH13750;
 XX DT 26-JUN-2001 (first entry)
 XX DE Human cDNA sequence SEQ ID NO:10656.
 XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX OS Homo sapiens.
 XX PN EP1074617-A2.
 XX PD 07-FEB-2001.
 XX PF 28-JUL-2000; 2000EP-0116126.
 XX PR 29-JUL-1999; 99JP-0248036.
 XX PR 27-AUG-1999; 99JP-0300253.
 XX PR 11-JAN-2000; 2000JP-0118776.
 XX PR 02-MAY-2000; 2000JP-0183767.
 XX PR 09-JUN-2000; 2000JP-0241899.
 XX PA (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX DR WPI: 2001-318749/34.
 XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX PS Claim 8; SEQ ID 10656; 2537pp + CD ROM; English.
 XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX SQ Sequence 1229 BP; 351 A; 243 C; 296 G; 339 T; 0 other;

alignment_scores:
 Quality: 372.00 Length: 120
 Ratio: 3.758 Gaps: 2
 Percent Similarity: 82.500 Percent Identity: 60.000

alignment_block:
 US-09-327-750D-34 x AAH13750 ..

Align seg 1/1 to: AAH13750 from: 1 to: 1229

```

1 MetAlaSerLysValLysGlnValIleLeuAspLeuThrValGluLysAs 17
||||| ||||| : : : : : ||||| : : : : :
191 ATGGAGTCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240

17 pLysLysAsnLysLysGlyGly...LysAlaSerLysGlnSerGluGlu 33
: : : : : : : : : : : : : : : : : : : : : : : : : : :
241 CCAACAAGAAACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 290

33 LuSerHisHisLeuGluValGluValGluValGluValGluValGlu 49
||||| ||||| : : : : : ||||| : : : : :
291 AATCCCGCCATTGGGAGGGGGTGAAGCCAGAGGAGGAGGAGGAGGAGGAG 340

50 ArgArg...LysValArgArgLeuValProAsnPheLeuTrpAlaIlePr 65
||||| ||||| : : : : : ||||| : : : : :
341 AGCGGGGGCGAGTATGAGCGAGTATGAGCGAGTATGAGCGAGTATGAG 390

65 OAsnArgHisValAspHisSerGluGlyGlyGluValGluValGluVal 82
||||| ||||| : : : : : ||||| : : : : :
391 TAATAGGCATATTGAGCAGACATGAGCGAGGAGGAGGAGGAGGAGGAGGAG 440

82 aLcGlyGlnValMetGluAlaLysArgHisSerLysGluGlnGlnMetArg 98
||||| ||||| : : : : : ||||| : : : : :
441 TAGGCAGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 490

99 ProTyrThrArgPheArgThrProGluProAspAsnHisTyrAspPheCy 115
||||| ||||| : : : : : ||||| : : : : :
491 CACTATATGCGCTTCCAACTCTGTGACCTGACCACTTATGACTTTTG 540

```

115 sleuillePro 118
|||||
541 CCTCATACCT 550

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.AAC85548

seq_documentation_block:

ID AAC85548 standard; cDNA; 1364 BP.

XX

AC AAC85548;

XX 04-JUN-2001 (first entry)

DE cDNA encoding CDIFF-4, Incyte ID No. 1990956B1.

XX Human; cell differentiation; CDIFF; agonist; antagonist; epilepsy;
KW cell proliferation; Alzheimer's disease; schizophrenic disorder;
KW arteriosclerosis; cancer; atherosclerosis; diabetes mellitus; ss.
XX Homo sapiens.

XX FH Key Location/Qualifiers
XX CDS 250..612
XX FT /*tag= a
XX FT /product= "CDIFF-4"

XX WO200119860-A2.

XX PN 22-MAR-2001.

XX PF 14-SEP-2000; 2000WO-US25435.

XX PR 15-SEP-1999; 99US-0154140.

XX PR 06-DEC-1999; 99US-0169155.

XX PA (INCY-) INCYTE GENOMICS INC.

XX PI Tang YF, Hillman JL, Yue H, Reddy R, Lal P, Shah P, Azimzai Y;
PI Baughn MR, Lu DAM, Bandman O, Shih LL, Patterson C;

XX WPI: 2001-211447/21.

DR P-PSDB: AAB47126.

XX Isolated polypeptides and polynucleotides involved in cell
PT differentiation are used for treatment, prevention and diagnosis of
PT cell proliferative, developmental and neurological disorders e.g.
PT cancer and Alzheimer's disease -

XX Claim 5; Page 121; 137pp; English.

XX The sequences given in AAC85545-72 encode human polypeptides involved
CC in cell differentiation (CDIFF). CDIFF polypeptides and agonists of
CC these are used to treat a disease or condition associated with
CC decreased expression of functional CDIFF. An antagonist of CDIFF is
CC used to treat a disease or condition associated with over expression
CC of functional CDIFF. CDIFF polypeptides may be used for the treatment,
CC prevention and diagnosis of cell proliferative, developmental and
CC neurological disorders, such as Alzheimer's disease, schizophrenic
CC disorders, arteriosclerosis, cancer, atherosclerosis, diabetes mellitus
CC and epilepsy. The CDIFF-4 sequence is homologous to Mus musculus
CC REX-3. This sequence maps to chromosome 1 within the interval from
CC 152.2 to 157.4 centiMorgans, to chromosome 3 within the interval from
CC 157.4 to 158.0 centiMorgans, and to the X chromosome within the interval
CC from 104.9 to 150.3 centiMorgans.

XX Sequence 1364 BP; 411 A; 269 C; 322 G; 362 T; 0 other;

alignment_scores:

Quality: 372.00 Length: 120
Ratio: 3.758 Gaps: 2
Percent Similarity: 82.500 Percent Identity: 60.000

alignment_block:

US-09-327-750D-34 x AAC85548

Align seq 1/1 to: AAC85548 from: 1 to: 1364

1 MetAlaSerLysValLysGlnValLleLeuAspLeuThrValGluLysAs 17
|||||
250 ATGGAGTCCAAAGAGAGAACTAGCGGCAACAATCTCAACGGGGAATGC 299
17 pLysLysAsnLysLysGlyGly...LysAlaSerLysGlnSerGluGluG 33
|||||
300 CCAACAAGAAACGAAGGAGGAGGAGCGGCCCCACGAGAATGAAGAAG 349
33 LuSerHisLysLeuGluValGluAsnLysLysProGlyGlyAsnVal 49
|||||
350 AATCCCGCCCATTTGGAGGGGTGAAGCCAGACCGCTGGAGGAATATC 399
50 ArgArg...LysValArgArgLeuValProAsnPheLeuTlpAlaIlePr 65
|||||
400 AGCGGGGCGAGTTAGGCGACTTGCTCTAATTTTCGATGGCCATACC 449
65 OAsnArgHisValAspHisSerGluGlyGlyValGluValGlyArgPheV 82
|||||
450 TAATAGGCATATTGAGCAATGAAGAGAGAGATGATGATAGAAAGTTTG 499
82 alGlyGlnValMetGluAlaLysArgHisSerLysGluGlnGlnMetArg 98
|||||
500 TAGGCAGATGATGGAATCAAGAGAGAGACTAGGGAACAGCAGATGAGG 549
99 ProTyrThrArgPheArgThrProGluProAsnHisTyrAspPheCy 115
|||||
550 CACTATATGCGCTTCAACACTCTGAACTGACACCAACATTATGACTTTG 599
115 sleuillePro 118
|||||
600 CCTCATACCT 609

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.AAF23529

seq_documentation_block:

ID AAF23529 standard; DNA; 891 BP.

XX

AC AAF23529;

XX

DT 22-MAR-2001 (first entry)

XX

DE Human NADE DNA.

XX

KW Neurotrophin receptor; p75-NTR; NGF-induced apoptosis;
KW neurogenetic disease; NF-kappaB; ds.

XX

OS Homo sapiens.

XX

PN WO200075278-A2.

XX

PD 14-DEC-2000.

XX

PF 07-JUN-2000; 2000WO-US15621.

XX

PR 07-JUN-1999; 99US-0327750.

XX

PA (UYCO) UNIV COLUMBIA NEW YORK.

XX

PI Sato T;

XX

DR WPI: 2001-061707/07.

XX

PT New p75-neurotrophin receptor-associated cell death executor (NADE) and
PT the gene encoding NADE, useful for modulating the activity of p75NTR
PT and for detecting neurodegenerative diseases -

PS Disclosure: Fig 1; 134pp; English.

486 CATATGGAATATTCATGGAGGATGAGAGAAATCAGAGAAACTTAG 535
93 sGluGlnMetArgProTyrThrArgPheArgThrProGluProAspA 110
|||||
536 GGAGCTCAGTTGAGGAATGTCTGGTATCTTATGGGGAGCTCTCTA 585
110 snHisTyrAsp.....PheCysLeuIlePro 118
|||||
586 ATCACCATGACCATCATGATGAATTTGCTTATGCCT 623

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAF23528

seq_documentation_block:

ID AAF23528 standard; DNA; 700 BP.

AC AAF23528;

DT 22-MAR-2001 (first entry)

XX Mouse NADE DNA.

XX Neurotrophin receptor; p75-NTR: NGF-induced apoptosis;
KW neurogenetic disease; NF-kappaB; ds.

XX Mus sp.

XX W0200075278-A2.

XX 14-DEC-2000.

XX 07-JUN-2000: 2000WO-US15621.

XX 07-JUN-1999: 99US-0327750.

XX (UYCO) UNIV COLUMBIA NEW YORK.

XX Sato T;

XX WPI: 2001-061707/07.

XX New p75-neurotrophin receptor-associated cell death executor (NADE) and
PT the gene encoding NADE, useful for modulating the activity of p75NTR
PT and for detecting neurodegenerative diseases -

XX Claim 12; Fig 1; 134pp; English.

XX The present invention relates to a purified polypeptide capable of
CC binding neurotrophin receptor (p75-NTR). The invention is useful for
CC NGF-induced apoptosis, which plays an important role in neurogenetic
CC diseases. The peptide of the invention and p75NTR are useful for
CC inhibiting NF-kappaB activation in a cell or a subject, for inducing
CC caspase-2 and caspase-3 activity to cleave poly (ADP-ribose) polymerase
CC and fragment nuclear DNA in a cell by co-expression of (I) and p75-NTR.

XX Sequence 700 BP; 177 A; 188 C; 203 G; 132 T; 0 other;

alignment_scores:

Quality: 144.00 Length: 129
Ratio: 2.000 Gaps: 3
Percent Similarity: 55.814 Percent Identity: 28.682

alignment_block:

US-09-327-750D-34 x AAF23528 ..

Align seg 1/1 to: AAF23528 from: 1 to: 700

18 LysLysAsnLysLysGlyGlyLysAlaSerLysGlnSerGluGluLuse 34

162 AAACAAATTCATCATGCGCCATGTCCACGAGAAACGAGAGACTGA 211

34 RHSHSHLeuGluGluValGluAsnLysLysProGlyGly..... 47

|||||
212 GCAGCCCTGCAGANTGGACAGGAAGACGCCCTGTGGGAGGAGGTGAGG 261
47
262 GCCACCAAGCTGCTGCAACAAACAACAACAACAACAACAACAACAACCA 311
48AsnValArgArgLysValArgLeuValProAsnPheLeuTr 62
312 AACCAACCAAGAGAGCGAGGCTGCCGACCTGCCCTAACTTCCGATG 361
62 pAlaIleProAsnArgHisValAspHisSerGluGly.....GlyGluG 77
|||||
362 GGCCATTCCCAACAGGAGATGAATGACGGTGTGGTGAGATGGAGATG 411
77 luValGlyArgPheValGlyGlnValMetGluAlaLysArgHisSerLys 93
|||||
412 ATATGGAATCTTCATGGAGGATGAGAGATCCGAGAAAGCTTAGG 461
94 GluGlnMetArgProTyrThrArgPheArgThrProGluProAspAs 110
|||||
462 GAGCTACAGCTGAGAAATGTCTACGCATCCTTATGGGGAGCTGTCTAA 511
110 nHisTyrAsp.....PheCysLeuIlePro 118
|||||
512 CCACCAAGATCACCATGATGAATTCGCCCTTATGCCT 548

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AA01004

seq_documentation_block:

ID AAC01004 standard; cDNA; 485 BP.

XX AAC01004;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 1002.

XX Humao; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI: 2000-500381/45.

XX P-PSDB; AAG00998.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1; SEQ ID 1002; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from

OS Homo sapiens.
 XX CN1296969-A.
 XX PD 30-MAY-2001.
 XX PF 23-NOV-1999; 99CN-0124078.
 XX PR 23-NOV-1999; 99CN-0124078.
 XX PA (SHAN-) SHANGHAI BORONG GENE DEV CO LTD.
 XX PI Mao Y, Xie Y;
 XX DR WPI: 2001-483897/53.
 XX DR P-PSDB; AAG66407.
 XX PT Polypeptide-human X chromosome linked gene expression protein 14 and
 PT polynucleotide for coding said polypeptide -
 XX Claim 6; Page 24 (Disclosure); 31pp; Chinese.
 CC The present sequence is the coding sequence for human X chromosome linked
 CC gene expression protein 14. The protein and coding sequence are useful
 CC for treating diseases e.g. cancer and HIV infection.
 XX Sequence 767 BP; 224 A; 143 C; 215 G; 185 T; 0 other;

alignment_scores:
 Quality: 111.00 Length: 132
 Ratio: 1.461 Gaps: 6
 Percent Similarity: 57.576 Percent Identity: 31.818

alignment_block:

US-09-327-750D-34 x AAH75810 ..
 Align seg 1/1 to: AAH75810 from: 1 to: 767
 1 MetalaSerLysValLysGlnValLleLeuAspLeuThrVal..... 14
 101 ATGGAGTCCCAAGAGGAGGAGCGTTAAACAATCTCATCGTGGAAATGT 150
 15GLuYAspLysLysAsnLysGlyGlyLysAlaSerLysG 29
 151 CAACGAGAAATGATGAAAGATGAAAGGAGCAAGTTCCTAATAA. 199
 29 InSerGluGluGluSerHisLysLeuGluValGluAsnLysLysPro 45
 200 ..GGGAGGCCCTTGGCCCTACCTTTGATGTTAGTGAATCTGTGTGCT 247
 46 GlyGlyAsnValArgArg.....LysValArgArgLeuValProAspPheLe 61
 248 AGAGAAACCGTAGGCGGTTCGCGTTAGGAGGCCCTCCTGCATATAG 297
 61 uTPAlaIleProAsnArgHisValasp.....HisSerG 73
 298 ATGGGACATATGATAGGCTTGGAGGCCACAGGAGGATGAGAGAGG 347
 73 LuGlyGlyGluGluValGlyArgPheValGlyLysValMetGluAlaLys 89
 348 ACATATGGAAGGATGGGAGGAGGTGAGACAGCTGATGGGAAAGCTG 397
 90 ArgHisSerLysGluGlnMetArgProTyrThrArgPheArgThrPr 106
 398 AGG.....GAAAGCAGTTGAGTCTCATGCTCGCGGAGCAGCAGC 438
 106 oGluPro.....AspAsnHisTyrAspPheCysLeuIlePro 118
 439 TGATCCCCCTCACCATGACCATCAGATGATTTTGCCTTATGCC 484

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI58581

seq_documentation_block:
 ID AAI58581 standard; cDNA; 862 BP.
 XX AC AAI58581;
 XX DT 22-OCT-2001 (first entry)
 XX DE Human polynucleotide SEQ ID NO 784.
 XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; seq
 XX OS Homo sapiens.
 XX PN WO200153312-A1.
 XX PD 26-JUL-2001.
 XX PF 26-DEC-2000; 2000WO-US34263.
 XX PR 21-JAN-2000; 2000US-0488725.
 XX PR 25-APR-2000; 2000US-0552317.
 XX PR 09-JUL-2000; 2000US-0598042.
 XX PR 19-JUL-2000; 2000US-0620312.
 XX PR 03-AUG-2000; 2000US-0653450.
 XX PR 14-SEP-2000; 2000US-0662191.
 XX PR 19-OCT-2000; 2000US-0691036.
 XX PR 29-NOV-2000; 2000US-0727344.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX WPI: 2001-442253/47.
 XX P-PSDB; AAM39425.
 XX PT Novel nucleic acids and polypeptides, useful for treating disorders
 XX such as central nervous system injuries -
 XX Claim 1; SEQ ID NO 784; 10078pp; English.

The invention relates to human nucleic acids (AAI57798-AAI61369) and the encoded polypeptides (AAM38642-AAM42213) with nootropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: immune system suppression, Actin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed specification.

Sequence 862 BP; 235 A; 183 C; 238 G; 206 T; 0 other;

alignment_scores:
 Quality: 111.00 Length: 132
 Ratio: 1.461 Gaps: 6
 Percent Similarity: 57.576 Percent Identity: 31.818

alignment_block:


```

XX SQ Sequence 662 BP; 177 A; 144 C; 192 G; 137 T; 12 other;
alignment_scores:
  Quality: 103.00 Length: 132
  Ratio: 1.411 Gaps: 6
  Percent Similarity: 55.303 Percent Identity: 30.303
alignment_block:
US-09-327-750D-34 x AAC03880
Align seg 1/1 to: AAC03880 from: 1 to: 662
1 MetAlaSerLysValLysGlnValIleLeuAspLeuThrVal..... 14
207 ATGGAGTCCAAAGAGGACGAGGCTTAACAATCTCATCTGTGGAATGT 256
15 .....GluLysAspLysLysAsnLysLysGlyGlyLysAlaSerLysG 29
257 CAACGAGGAAATGATGAAAAGATGAAAAGGAGCAAGTGTCTAATAAA. 305
29 InSerGluGluGluSerHisHisLeuGluGluValGluAsnLysLysPro 45
306 ..GGGAGCCCTTGGCCCTACCTTTGATGCTGTGTGTAATACGTGTGCT 353
46 GlyGlyAsnValArgArg...LysValArgArgLeuValProAsnPhel 61
354 AGAGGAATCTAGCGGCTTCCGCTTAGCAGCCCTCTGCAGTATAG 403
61 uTrpAlaIleProAsnArgHisValAsp.....HisSerG 73
404 ATGGGATATGATGATAGGCTTGGAGAACCCACAGGCAAGGATGANAGA 453
73 LuGlyGlyGluGluValGlyArgPheValGlyGlnValMetGluAlaLys 89
454 AGAATATGGAAGGATTTGGGAGGAGGTGAGACAGCTGATGGAAGAAG 503
90 ArgHisSerLysGluGlnGlnMetArgProTyrThrArgPheArgThrPr 106
504 AGG.....GAAAAGCAGTTGAGTCACTAGTCTGCGGGCAGTCAGCAC 544
106 oGluPro.....AspAsnHisTyrAspPheCysLeuIlePro 118
545 TGACCCCTCCACCATGACCATCATGATGATGATGTTTCNNWATGCC 590
seq_name: /SDS2/gcgdata/geneseq/geneseq/NA2000.DAT: AAC03879
seq_documentation_block:
AAC03879 standard; cdna; 698 BP.
AAC03879;
06-OCT-2000 (first entry)
XX Human secreted protein 5' EST, SEQ ID NO: 3877.
XX Human; 5' EST; expressed sequence tag; secreted protein; cdna isolation;
XX gene therapy; chromosome mapping; ss.
XX Homo sapiens.
XX EP1033401-A2.
XX 06-SEP-2000.
XX 21-FEB-2000; 2000EP-0200610.
XX 26-FEB-1999; 99US-0122487.
XX (GEST) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI

```

```

XX DR WPI; 2000-500381/45.
XX P-PSDB; AAG03873.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
PS Claim 1; SEQ ID 3877; 71pp + CD-ROM; English.
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. An ORF has been identified within the
XX sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
XX derived from 30 different tissues. EST sequences usually correspond
XX mainly to the 3' untranslated region (UTR) of the mRNA because they are
XX often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
XX well suited for isolating cDNA sequences derived from the 5' ends of
XX mRNAs and even in those cases where longer cDNA sequences have been
XX obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
XX cDNAs with intact 5' ends and can therefore be used to obtain full length
XX cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
XX gene therapy and chromosome mapping procedures. They are used to obtain
XX upstream regulatory sequences and to design expression and secretion
XX vectors.
XX SQ Sequence 698 BP; 187 A; 151 C; 210 G; 139 T; 11 other;
alignment_scores:
  Quality: 103.00 Length: 132
  Ratio: 1.411 Gaps: 6
  Percent Similarity: 55.303 Percent Identity: 30.303
alignment_block:
US-09-327-750D-34 x AAC03879
Align seg 1/1 to: AAC03879 from: 1 to: 698
1 MetAlaSerLysValLysGlnValIleLeuAspLeuThrVal..... 14
243 ATGGAGTCCAAAGAGGACGAGGCTTAACAATCTCATCTGTGGAATGT 292
15 .....GluLysAspLysLysAsnLysLysGlyGlyLysAlaSerLysG 29
293 CAACGAGGAAATGATGAAAAGATGAAAAGGAGCAAGTGTGCTAATAAAA. 341
29 InSerGluGluGluSerHisHisLeuGluGluValGluAsnLysLysPro 45
342 ..GGGAGCCCTTGGCCCTACCTTTGATGCTGTGTGTAATACGTGTGCT 389
46 GlyGlyAsnValArgArg...LysValArgArgLeuValProAsnPhel 61
390 AGAGGAATCTAGCGGCTTCCGCTTAGCAGCCCTCTGCAGTATAG 439
61 uTrpAlaIleProAsnArgHisValAsp.....HisSerG 73
440 ATGGGATATGATGATAGGCTTGGAGAACCCACAGGCAAGGATGANAGAAG 489
73 LuGlyGlyGluGluValGlyArgPheValGlyGlnValMetGluAlaLys 89
490 AGAATATGGAAGGATTTGGGAGGAGGTGAGACAGCTGATGGAAGAAGCTG 539
90 ArgHisSerLysGluGlnGlnMetArgProTyrThrArgPheArgThrPr 106
540 AGG.....GAAAAGCAGTTGAGTCACTAGTCTGCGGGCAGTCAGCAC 580
106 oGluPro.....AspAsnHisTyrAspPheCysLeuIlePro 118
581 TGACCCCTCCACCATGACCATCATGATGATGATGTTTCNNWATGCC 626
seq_name: /SDS2/gcgdata/geneseq/geneseq/NA2001.DAT: AAI60367
seq_documentation_block:

```


OM of: US-09-327-750D-34 to: GenEmbl.* out_format : pfs
 Date: Mar 11, 2002 3:34 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=framet.p2n.model -DEV=xlh
 -O=/sgn2.1/USPTO_spool/US09327750/runat_11032002_101154_20319/app_query.fasta_1.1472
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000 -QGAPOP=4.500
 -CGAPEXT=0.050 -XGAPOP=10.000 -YGAPEXT=0.500 -FGAPOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blomsum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -OUTFMT=pfs -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZ=500
 -MINLEN=0 -MAXLEN=2000000000 -USER=US09327750_0CGN1.1_8673
 -NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
 -NO_XLPXY -WAIT_THREADS=1

Search information block:
 Query: US-09-327-750D-34
 Query length: 118
 Database: GenEmbl.*
 Database sequences: 1472140
 Database length: 34134837
 Search time (sec): 4557.230000

score list:	Strid	Orig	ZScore	EScore	Len	Documentation
gb.pr:AK000959	+	372.00	661.44	1.48-28	1229	AK000959 Homo sapiens CDNA FL3
gb.pr:AL100231	+	372.00	660.55	1.6e-28	1364	AL100231 Sequence 32 from Pat
gb.pr:HS035319	+	372.00	627.19	1.2e-26	6948	AL035494 Human DNA sequence
gb.pr:HUM06C	+	171.50	302.23	1.5e-08	891	M3188 Human unknown protein f
gb.pr:AF187064	+	171.50	302.23	1.5e-08	891	AF187064 Homo sapiens p75NTR-as
gb.pr:BC003190	+	170.00	300.51	1.8e-08	793	BC003190 Homo sapiens, p75NTR-a
gb.pr:AF187065	+	148.00	264.39	1.9e-06	519	AF187065 Rattus norvegicus p75N
gb.sts:G72708	+	145.00	260.75	3.0e-06	421	G72708 MARC 4953-4954:99193031
gb.pr:AF187066	+	145.00	256.44	5.2e-06	700	AF187066 Mus musculus p75NTR-as
gb.pr:AF097440	+	145.00	254.75	6.5e-06	854	AF097440 Mus musculus brain exp
gb.pr:HSV351F8	+	138.00	208.36	0.0025	45678	AF097440 Mus musculus brain exp
gb.sts:G55294	+	118.50	211.86	0.0016	477	G55294 human STS SHGC-37409, se
gb.pr:AF097439	+	116.50	204.02	0.0044	785	AF097439 Mus musculus brain exp
gb.pr:AF183416	+	112.50	198.50	0.0088	642	AF183416 Homo sapiens ovarian g
gb.pr:AF220189	+	112.50	196.35	0.0117	828	AF220189 Homo sapiens uncharact
gb.pr:HS1988P4	+	112.50	164.41	0.6998	35714	AL008708 Human DNA sequence
gb.pr:AF251053	+	111.00	194.55	0.0147	744	AF251053 Homo sapiens X-linked
gb.pr:AF208272	+	111.00	192.95	0.0180	898	AX078272 Sequence 76 from Paten
gb.pr:HSV870H8	+	111.00	162.82	0.8586	31321	AF078272 Sequence 76 from Paten
gb.pr:AL133348	+	111.00	160.62	1.14	40384	AL133348 Human DNA sequence
gb.pr:HS080B1	+	111.00	160.53	1.15	41029	AL022169 Homo sapiens chromos
gb.pr:AF237783	+	109.50	191.32	0.0222	791	AF237783 Homo sapiens brain-exp
gb.pr:AF097438	+	105.00	182.74	0.0668	835	AF097438 Mus musculus brain exp
gb.pr:AF051347	+	105.00	182.71	0.0670	838	AF051347 Mus musculus REX-3 mRN
gb.pr:AF097437	+	105.00	174.26	0.1981	2269	AF097437 Mus musculus Bex1 pr
gb.pr:HS714B7	+	102.50	137.78	21.32	98274	AF29755 Human DNA sequence
gb.pr:AF071813	+	96.00	163.24	0.8138	1225	AF071813 Streptococcus pneumon
gb.pr:AC023561	+	95.50	120.50	195.44	169808	AC023561 Homo sapiens chromo
gb.pr:AC074348	+	95.00	120.24	202.20	157511	AC074348 Homo sapiens clone
gb.pr:AC048344	+	95.00	119.13	232.92	179380	AC048344 Homo sapiens chromo
gb.pr:AC023341	+	94.50	118.72	245.66	169365	AC023341 Homo sapiens chromo
gb.pr:HSAC000373	+	94.00	130.14	56.75	39602	AC000373 Human cosmid g1862d2
gb.pr:HSAC000367	+	94.00	129.38	62.62	43349	AC000367 Human Cosmid g1862d2
gb.pr:AC012232	+	94.00	117.57	284.63	173337	AC012232 Homo sapiens clone
gb.pr:HSV353P9	+	92.50	118.28	259.95	116561	AL023653 Human DNA sequence
gb.pr:AL1355809	+	92.50	118.01	269.13	120340	AL1355809 Human DNA sequence
gb.pr:AF195953	+	92.50	113.34	489.70	208618	AF195953 Homo sapiens membra
gb.in:CSA7431	+	92.00	151.74	3.56	2029	AJ007431 Cupienius saiei mRNA
gb.pr:AL512842	+	92.00	114.96	397.64	154889	AL512842 Homo sapiens chromo
gb.pr:AC091457	+	92.00	112.30	559.73	212080	AC091457 Mus musculus clone

gb.htg:AL139096 + 91.50 115.47 372.68 131198 I AL139096 Homo sapiens chr
 gb.pr:U72787 + 91.00 131.18 49.69 18510 I U72787 Homo sapiens cosmid
 gb.pr:AL359643 + 91.00 112.53 543.53 166863 I AL359643 Human DNA sequen
 gb.pr:HS078N14 - 91.00 111.78 597.92 182150 I AL109654 Human DNA sequen
 gb.htg:B1011H02 + 90.50 119.60 219.49 65190 I AL442109 Oryza sativa chro

seq_name: gb.pr:AK000959

seq_documentation_block:
 LOCUS AK000959 1229 bp mRNA PRI 22-FEB-2000
 DEFINITION Homo sapiens CDNA FLJ10097 fis, clone HEMBA1002458, weakly similar
 to OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74.
 ACCESSION AK000959
 VERSION AK000959.1 GI:7021945
 KEYWORDS oligo clapping; fis (full insert sequence).
 SOURCE Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to
 mRNA, clone lib:HEMBAL clone:HEMBAL1002458.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (sites)
 AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
 Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H.,
 Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K.,
 Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
 Wakamatsu,A., Nakamura,Y., Nagahara,K., Masuho,Y. and Oshima,A.
 NEDO human cDNA sequencing project
 Unpublished (2000)

JOURNAL
 REFERENCE 2 (bases 1 to 1229)
 AUTHORS Isogai,T. and Otsuki,T.
 TITLE Direct Submission

JOURNAL
 Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao

Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana,
 Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
 Tel:81-438-52-3951, Fax:81-438-52-3952)

COMMENT
 NEDO human cDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan; cDNA full insert
 sequencing: Research Association for Biotechnology; cDNA library
 construction, 5'- & 3'-end one pass sequencing and clone selection:
 Helix Research Institute (supported by Japan Key Technology Center
 etc.) and Department of Virology, Institute of Medical Science,
 University of Tokyo.

FEATURES
 Location/Qualifiers
 1..1229
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HEMBAL1002458"
 /clone_lib="HEMBAL"
 /dev_stage="embryo, 10 weeks"
 /tissue_type="whole embryo, mainly head"
 /note="cloning vector: pME18SFL3"
 191..553
 /codon_start=1
 /cotton_start=1
 /protein_id="BAA91443.1"
 /db_xref="gi:7021946"
 /translation="MESKEELANLNNGENAEQEGGQAPTONEEERHLGGEGG
 KPGGNIRGRVRRVLPNFRWPAIPNRHIEHNEARDVDFVFGQWMEIKRKRTRQQRHRY
 MRQPEPDNDHYDCLIP"

BASE COUNT 351 a 243 c 296 g 339 t
 ORIGIN

alignment_scores:
 Quality: 372.00 Length: 120
 Ratio: 3.758 Gaps: 2
 Percent Similarity: 82.500 Percent Identity: 60.000

alignment_block:
 US-09-327-750D-34 x AK000959
 Align seg 1/1 to: AK000959 from: 1 to: 1229

details see <http://bacpac.med.buffalo.edu/VECTOR: pcYPAC2>
 IMPORTANT: This sequence is not the entire insert of clone 635G19.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we arrange for a small overlap between
 neighbouring submissions.
 The true left end of clone U65A4 (281014) is at 69549 in this
 sequence. The true right end of clone U101D3 (285997) is at 100 in
 this sequence.

FEATURES

Location/Qualifiers

source

```

1..69648
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q22.1-22.3"
/clone="RP4-635G19"
/clone.lib="RCI-4"
339..483
/feature="TAR1 repeat: matches 707..749 of consensus"
501..561
/feature="AluJb repeat: matches 137..311 of consensus"
662..897
/feature="118 copies 2 mer aa 70% conserved"
915..1331
/feature="WSTA repeat: matches 1..419 of consensus"
1334..1517
/feature="92 copies 2 mer ga 77% conserved"
1927..2020
/feature="MIR repeat: matches 109..217 of consensus"
2066..2262
/feature="L1ME2 repeat: matches 5825..6022 of consensus"
2263..2574
/feature="AluSp repeat: matches 1..313 of consensus"
2575..2819
/feature="L1ME2 repeat: matches 5583..5825 of consensus"
2821..3039
/feature="Alu repeat: matches 1..311 of consensus"
3104..3384
/feature="L1ME2 repeat: matches 5304..5591 of consensus"
3391..4007
/feature="L1MB1 repeat: matches 5558..6162 of consensus"
4012..4137
/feature="Trigger2a repeat: matches 11..123 of consensus"
4141..4252
/feature="L1 repeat: matches 2920..3031 of consensus"
4265..4350
/feature="HY4 repeat: matches 1..87 of consensus"
4350..4678
/feature="MER74A repeat: matches 28..369 of consensus"
4914..5095
/feature="MER3A repeat: matches 8..189 of consensus"
5224..5555
/feature="L2 repeat: matches 2423..2746 of consensus"
6040..6086
/feature="L2 repeat: matches 1676..1720 of consensus"
7413..7478
/feature="Alu repeat: matches 239..304 of consensus"
7628..7928
/feature="AluB8 repeat: matches 1..316 of consensus"
7952..7987
/feature="18 copies 2 mer aa 81% conserved"
7990..8305
/feature="AluJc repeat: matches 1..309 of consensus"
8436..8538
/feature="MIR repeat: matches 99..212 of consensus"
10836..11442
/feature="L1MD repeat: matches 1357..1972 of consensus"
11443..11571
/feature="FLAN_A repeat: matches 2..129 of consensus"
11572..11946
/feature="L1MD repeat: matches 977..1357 of consensus"
11947..12245
/feature="AluX repeat: matches 1..301 of consensus"
12246..13098

```

```

repeat_region
/feature="L1MD repeat: matches 275..977 of consensus"
13094..13223
/feature="L1MD repeat: matches 17..113 of consensus"
13523..13820
/feature="L1MD repeat: matches 177..505 of consensus"
13823..13932
/feature="55 copies 2 mer ct 77% conserved"
13936..13977
/feature="L1MD repeat: matches 201..241 of consensus"
13978..14240
/feature="AluSg1 repeat: matches 1..270 of consensus"
14245..14534
/feature="AluSc repeat: matches 1..290 of consensus"
14535..14721
/feature="L1MD repeat: matches 2..201 of consensus"
14724..15604
/feature="L1MC/D repeat: matches 4646..5565 of consensus"
15605..15909
/feature="AluX repeat: matches 24..312 of consensus"
16069..16363
/feature="AluX repeat: matches 1..291 of consensus"
16482..16641
/feature="AluJb repeat: matches 136..295 of consensus"
16671..16953
/feature="AluX repeat: matches 1..284 of consensus"
17164..17280
/feature="L1P repeat: matches 4499..4619 of consensus"
17326..17377
/feature="26 copies 2 mer at 79% conserved"
17704..17863
/feature="AluSc repeat: matches 118..295 of consensus"
17864..18175
/feature="AluB8 repeat: matches 1..318 of consensus"
18176..18298
/feature="AluSc repeat: matches 1..118 of consensus"
18848..18873
/feature="13 copies 2 mer tg 100% conserved"
18876..19171
/feature="AluSg repeat: matches 1..295 of consensus"
19183..19232
/feature="25 copies 2 mer ac 100% conserved"
19246..19533
/feature="AluY repeat: matches 5..292 of consensus"
19539..19837
/feature="MER74B repeat: matches 61..368 of consensus"
20016..20194
/feature="MER74A repeat: matches 295..487 of consensus"
20394..20692
/feature="AluSg repeat: matches 1..302 of consensus"
20701..20809
/feature="L1R1 repeat: matches 682..785 of consensus"
20877..21061
/feature="L1R1 repeat: matches 1..195 of consensus"
21186..21636
/feature="L1R29 repeat: matches 3..498 of consensus"
21609..21737
/feature="MER34 repeat: matches 409..543 of consensus"
21771..22076
/feature="AluY repeat: matches 2..308 of consensus"
22878..22974
/feature="MIR repeat: matches 92..191 of consensus"
23452..23582
/feature="FLAN_C repeat: matches 1..131 of consensus"
23583..23624
/feature="21 copies 2 mer ta 76% conserved"
23735..24041
/feature="AluSg1 repeat: matches 1..308 of consensus"
24045..24072
/feature="14 copies 2 mer tt 89% conserved"
25274..25582
/feature="AluJb repeat: matches 1..289 of consensus"
25583..25747
/feature="match: GSS B63622"
misc_feature

```

repeat_region 25959..26012
/note="HERVPH21 repeat: matches 23. .79 of consensus"
26144..26456
/note="Alu repeat: matches 1. .311 of consensus"
repeat_region 27281..27432
/note="Alu repeat: matches 137. .289 of consensus"
27436..27507
/note="L1M4 repeat: matches 5360. .5434 of consensus"
repeat_region 27536..27600
/note="L1M4 repeat: matches 2220. .2283 of consensus"
repeat_region 27601..27975
/note="MLT1A1 repeat: matches 3. .365 of consensus"
repeat_region 27976..28273

alignment_scores:
Quality: 372.00 Length: 120
Ratio: 3.758 Gaps: 2
Percent Similarity: 82.500 Percent Identity: 60.000

alignment_block:
US-09-327-750D-34 x HS635G19

Align seg 1/1 to: HS635G19 from: 1 to: 69648

1 MetalSerLysValLysGlnValLeuAspLeuThrValGluLysAs 17
64293 ATGGAGTCAAGAGGAACTAGCGGCAACAACTCTCAACGGGGAATGC 64342
17 pLysLysAsnLysLysGlyGly...LysAlaSerLysGlnSerGluGluG 33
64343 CCAACAAGAAACGAAGAGGGGAGCGAGCGCCCGCCAGCAATGAAGAAG 64392
33 LuSerHisHisLeuGluValGluAsnLysLysProGlyGlyAsnVal 49
64393 AATCCCGCCATTGGAGGGGGTGAAGCCAGAGGCTGGAGGAATATC 64442
50 ArgArg...LysValArgValLeuValProAsnPheLeuTrpAlaIlePr 65
64443 AGCGGGGGCGAGTTAGCGGACTGTCCCTAATTTTCGATGGCCATACC 64492
65 OAsnArgHisValAspHisSerGluGlyGlyGluValGluValArgPheV 82
64493 TAATAGGCATATTGAGCACAATGAGCGAGAGATGATGAGAAAGTTTG 64542
82 alGlyGlnValMetGluAlaLysArgHisSerLysGluGlnMetArg 98
64543 TAGGCGCATGATGGAATCAAGAGAAAGACTAGGGAACAGCAGATGAGG 64592
99 ProTyrThrArgPheArgThrProGluProAspAsnHistyAspPheCy 115
64593 CACTATATGCGCTTCCAAACTCTGAACTGACACCACTTATGACTTTTG 64642
115 sleullePro 118
64643 CCTCATACCT 64652

seq_name: gb_pr: HUMOGC

seq_documentation_block:
LOCUS HUMOGC 891 bp mRNA PRI 07-MAR-1995
DEFINITION Human unknown protein from clone pHR74 mRNA, complete cds.
ACCESSION M38188 X56942
VERSION M38188.1 GI:189378
KEYWORDS
SOURCE Human ovarian granulosa cell line, cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 891)
Rapp.G., Freudenstein,J., Klaudiny,J., Mucha,J., Wempe,F.,
Zimmer,M. and Scheit,K.H.
Characterization of three abundant mRNAs from human ovarian
granulosa cells

JOURNAL DNA Cell Biol. 9 (7), 479-485 (1990)
MEDLINE 91025550
COMMENT Draft entry and computer-readable sequence for [DNA 9, 479-485
(1990)] kindly submitted
by K.H.Scheit, 27-AUG-1990.

FEATURES
source
1. .891
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="pHR74"
/cell_type="granulosa"
/tissue_type="ovary"
c...891
/note="protein of unknown function"
312. .647
/note="protein of unknown function"
/codon_start=1
/protein_id="AAA63232.1"
/db_xref="GI:189379"
/translation="NANTHOENEMEOPMONGEDRPLGGEGHQPAGNRRRGOARRLA
PHERAIPNRIQINDGCGGDDMEIFMEERIRKRLRELQRLNCLRLANGELSNHHD
HDEFCLEP"
361. .534
/note="protein of unknown function"
/codon_start=1
/protein_id="AAA63233.1"
/db_xref="GI:189380"
/translation="MERKTALEEVKATSLQEIQDGLADLPLIFDGPYPGRSMHMGW
VEMEWIKYSWRR"
251 a 182 c 224 g 234 t

BASE COUNT
ORIGIN

alignment_scores:
Quality: 171.50 Length: 126
Ratio: 2.257 Gaps: 3
Percent Similarity: 60.317 Percent Identity: 32.540

alignment_block:
US-09-327-750D-34 x HUMOGC

Align seg 1/1 to: HUMOGC from: 1 to: 891

12 LeuThrValGluLysAspLysLysGlyGlyLysLysAlaSerLy 28
267 TTACTTGTGTTCAAGAAAAACACCAAGAAAAAATCTCATATGC 316
28 sGlnSerGluGluGluSerHisHisLeuGluGluVal..... 40
317 AAATATTCCAGGAAAAACGAAGAGATGGAGCGCCTATCGAAGTGGAG 366
41GluAsnLysLysProGlyGlyAsn 48
367 AGGAAGACCGCCCTTTGGAGAGGTGAAGCCACAGCCTGCAGGAAT 416
49 ValArgArgLysValArgLeuValProAsnPheLeuTrpAlaIlePr 65
417 CGACGGGACAGCGCTCGCGGACTTGCCTTAATTTTCGATGGCCATACC 466
65 oAsnArgHisValAspHisSerGluGly.....GlyGluGluValGly 80
467 CAATAGGAGATCAATGATGGGATGGGATGGAGATGATGATGATAA 516
80 rgPheValGlyGlnValMetGluAlaLysArgHisSerLysGluGln 96
517 TATTTCATGGAGAGATGAGAAATACAGAAACAACTTAGGAGCTGCAG 566
97 MetArgProTyrThrArgPheArgThrProGluProAspAsnHisty 113
567 TTGAGGAATGTCCTCGTATCTTATGGGGAGCTCTCTAATCACCATGA 616
113 p.....PheCysLeuIlePro 118
|||||


```

|||||
162 AACAAATCTCATGCGCAATGTCACACAGGAAACGAAGATGGA 211
34 rHisHisLeuGluValGluAsnLysProGlyGly..... 47
212 GCAGCCCTGCAGAAATGGACAGGAGACCGCCCTGTGGGAGGTGAGG 261
47 ..... 47

262 GCCACGAGCTGCTGCAACAAACAAACAAACCAACCAATACACAC 311
48 .....AsnValArgArgLysValArgLeuValProAsnPheLeuTr 62
312 AACACACGAGAGCGGCGGCTGCGCAGCTTGCCTTAACCTCCGATG 361
62 palatleProAsnArgHisValAspHisSerGluGly.....GlyGluG 77
362 GGCCATTCACACAGGAGATGATGACGGGTTGGTGAGATGGAGATG 411
77 luValGlyArgPheValGlyGlnValMetGluAlaLysArgHisSerLys 93
412 ATATGGAAATGTTATGGAGGAGATGAGAGATCCGGAGAAAGCTTAGG 461
94 GluGlnMetArgProTyThrArgPheArgThrProGluProAspAs 110
462 GAGCTACAGCTGAGAAATGTTCTACGCATCTTATGGGGAGCTGTCTAA 511

110 nHisTyAsp.....PheCysLeuLeuPro 118
512 CCACCGAGTACCATGATGATGATTCGCTTATGCCT 548

seq_name: gb_ro:AF097440
seq_documentation_block:
LOCUS AF097440 854 bp mRNA ROD 13-APR-1999
DEFINITION Mus musculus brain expressed x-linked protein 3 (Bex3) mRNA,
complete cds.
ACCESSION AF097440
VERSION AF097440.1 GI:4580593
KEYWORDS
SOURCE house mouse.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 854)
Brown, A.L. and Kay, G.F.
Bex3, a gene with increased expression in parthenogenetic embryos,
is a member of a novel gene family on the mouse X chromosome
Hum. Mol. Genet. 8 (4), 611-619 (1999)
JOURNAL 99172070
MEDLINE
REMARK Erratum: [[published erratum appears in Hum Mol Genet 1999
May;8(5):943]]
REFERENCE 2 (bases 1 to 854)
AUTHORS Brown, A.L. and Kay, G.F.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1998) Cancer Unit, Queensland Institute of
Medical Research, Herston Rd, Brisbane, Qld 4029, Australia
FEATURES
SOURCE
1. 854
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/db_xref="dbEST:AA272375"
/chromosome="X"
/map="near Plp"
/tissue_type="pooled organs"
1. 854
/gene="Bex3"
172. 546
/gene="Bex3"
/codon_start=1
/product="brain expressed X-linked protein 3"
/protein_id="AAD24431.1"

```

```

/db_xref="GI:4580594"
/translation="MANVHOENEMEEOPLQNGQEDRPVGGEGHQPAANNNNNNHHN
NHHRRGOARRLAPFRWAIQNMNDGLGDDMDMFMEMREIRRLKRLQLRNC
LRLIMGELSNHHDHDEFLMP"
BASE COUNT 237 a 212 c 228 g 177 t
ORIGIN

alignment_scores:
Quality: 145.00 Length: 129
Ratio: 2.014 Gaps: 3
Percent Similarity: 55.814 Percent Identity: 28.682

alignment_block:
US-09-327-750D-34 x AF097440 ..
Align seg 1/1 to: AF097440 from: 1 to: 854

18 LysLysAsnLysLysGlyGlyLysAlaSerLysGlnSerGluGluGluSe 34
157 AAACAATCTCATGCGCAATGTCACACAGGAAACGAAGATGGA 206
34 rHisHisLeuGluValGluAsnLysProGlyGly..... 47
207 GCAGCCCTGCAGAAATGGACAGGAGACCGCCCTGTGGGAGGTGAGG 256
47 ..... 47

257 GCCACGAGCTGCTGCAACAAACAAACAAACCAACCAACCAATACAC 306
48 .....AsnValArgArgLysValArgLeuValProAsnPheLeuTr 62
307 AACACACGAGAGCGGCGGCTGCGCAGCTTGCCTTAACCTCCGATG 356
62 palatleProAsnArgHisValAspHisSerGluGly.....GlyGluG 77
357 GGCCATTCACACAGGAGATGATGACGGGTTGGTGAGATGGAGATG 406
77 luValGlyArgPheValGlyGlnValMetGluAlaLysArgHisSerLys 93
407 ATATGGAAATGTTATGGAGGAGATGAGAGATCCGGAGAAAGCTTAGG 456
94 GluGlnMetArgProTyThrArgPheArgThrProGluProAspAs 110
457 GAGCTACAGCTGAGAAATGTTCTACGCATCTTATGGGGAGCTGTCTAA 506
110 nHisTyAsp.....PheCysLeuLeuPro 118
507 CCACCGAGTACCATGATGATGATTCGCTTATGCCT 543

seq_name: gb_pr:HSV351F8
seq_documentation_block:
LOCUS HSV351F8 45678 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from cosmid V351F8, between markers DXS366 and
DXS87 on chromosome X contains ESTs.
ACCESSION 270719
VERSION 270719.1 GI:1261915
KEYWORDS x.
SOURCE human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 45678)
Whiteley, M.
Direct Submission
Submitted (09-APR-1995) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RO, UK. E-mail enquiries: humquery@sanger.ac.uk
IMPORTANT: This sequence is the entire insert of clone V351F8. The
true left end of clone V351F8 is at 1 in this sequence. The true
right end of clone V351F8 is at 45678.
V351F8 is from the human chromosome X-specific cosmid library.
Location/Qualifiers
FEATURES

```

```

source
1. 45678
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone="GHC-351F8"
/clone_lib="SCCV"
1. .70
repeat_region
/partial
/note="Alu repeat: matches 80. .1 of consensus"
642. .930
repeat_region
/note="Alu repeat: matches 1. .308 of consensus"
1406. .5179
repeat_region
/note="L1 element fragment"
5243. .5276
repeat_region
/note="17 copies of 2 mer 82 & conserved"
5279. .5567
/partial
/note="Alu repeat: matches 304. .1 of consensus"
9375. .10369
repeat_region
/note="MSTA element fragment"
10399. .12017
repeat_region
/note="MSTAR element fragment"
11318. .11978
repeat_region
/note="THEIR element fragment"
11715. .11846
repeat_region
/note="MLTIR element fragment"
12074. .12170
repeat_region
/note="MSTA element fragment"
12227. .12486
repeat_region
/note="MSTA element fragment"
12437. .12502
repeat_region
/note="MSTC element fragment"
12756. .13059
/partial
/note="Alu repeat: matches 308. .1 of consensus"
14478. .14650
repeat_region
/note="THEIR element fragment"
14491. .14536
repeat_region
/note="MSTA element fragment"
14715. .14831
repeat_region
/note="THEIR element fragment"
15099. .15533
repeat_region
/note="L1 element fragment"
15534. .15626
repeat_region
/note="31 copies of 3 mer 85 & conserved"
15566. .15619
repeat_region
/note="3 copies of 18 mer 98 & conserved"
16352. .16436
repeat_region
/note="L1 element fragment"
16728. .16763
repeat_region
/note="9 copies of 4 mer 94 & conserved"
16764. .17054
/partial
/note="Alu repeat: matches 308. .1 of consensus"
19420. .19449
repeat_region
/note="15 copies of 2 mer 87 & conserved"
complement(19537. .20013)
/misc_feature
/note="match: 3' EST N51315 clone 283089"
20071. .20114
repeat_region
/note="22 copies of 2 mer 98 & conserved"
20073. .20112
repeat_region
/note="10 copies of 4 mer 100 & conserved"
20073. .20126
repeat_region
/note="3 copies of 18 mer 87 & conserved"
23088. .23300
repeat_region
/note="L1 element fragment"
23311. .23490
repeat_region
/note="MLTIA element fragment"
23318. .23489
repeat_region
/note="MLTIB element fragment"
23362. .23491
repeat_region
/note="MLTIC element fragment"

repeat_region
23507. .24407
/note="L1 element fragment"
24376. .24662
repeat_region
/note="MSTC element fragment"
24379. .24491
repeat_region
/note="MSTA element fragment"
24688. .24738
repeat_region
/note="MSTA element fragment"
24810. .25016
repeat_region
/note="L1 element fragment"
25254. .25421
repeat_region
/note="L1 element fragment"
25439. .25696
repeat_region
/note="MER25 element fragment"
26284. .26343
repeat_region
/note="MLTID element fragment"
26428. .26491
repeat_region
/note="MLTID element fragment"
26465. .26518
repeat_region
/note="MLTIE element fragment"
27446. .27667
repeat_region
/note="L1 element fragment"
28954. .29247
/partial
/note="Alu repeat: matches 308. .1 of consensus"
30160. .30325
repeat_region
/note="2 copies of 83 mer 98 & conserved"
32588. .32875
/misc_feature
/note="match: 5' EST H68599 clone 239077"
complement(32825. .33230)
/misc_feature
/note="match: 3' EST H68239 clone 289077"
35551. .35802
repeat_region
/partial
/note="Alu repeat: matches 1. .260 of consensus"
35804. .35839
repeat_region
/note="18 copies of 2 mer 83 & conserved"
38625. .38705
repeat_region
/note="MLTIR element fragment"
38747. .38839
repeat_region
/note="MLTIR element fragment"
38949. .39032
repeat_region
/note="MLTIE element fragment"
38965. .39076
repeat_region
/note="MLTID element fragment"
BASE COUNT 14260 a 9135 c 9327 g 12956 t
ORIGIN

alignment_scores:
Quality: 138.00 Length: 127
Ratio: 2.000 Gaps: 3
Percent Similarity: 54.331 Percent Identity: 29.921

alignment_block:
US-09-327-750D-34 x HSV351F8
Align seg 1/1 to: HSV351F8 from: 1 to: 45678
12 LeuThrValGluLysAspLysLysLysLysLysGlyGlyLysAlaSerly 28
19408 TTATCTTCTAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 19457
28 sGlnSer.....Glug 32
19458 GGAATATGTCCCAAGGAAACAANGTTGTGGAGAGGCCCGCCAGTCGAGA 19507
32 luGluSerHisHisLysGluGluValGluAsnLysLysProGlyGlyAsn 48
19508 ATGAGCCCGCGCTTTAGGAGGTGTGAATACACAGGAGCGCTGGAGAAAT 19557
49 ValArgArgLysValArgArgLeuValProAsnPheLeuTrpAlaIlePr 65
19558 GTTAAAGGGGTTTGGGCTCCACCTGCCCGGGGTTTTCGAGAGGATGTGCC 19607

```

```

65  oasnaRgHisValasPhisSerGlu.....GlyGlyGluGluV 78
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
19608  CAATAGGCTTCGATACATGATGATAGATGAGATGAGATGATA 19657

78  alGlyArgPheValGlyClnValMetGluAlaLysArgHisSerLysGlu 94
::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||
19658  TGAACGGTTTCAGAGAGATGAGAGAGTAAGGAGGAAATAGGGAA 19707

95  GlnGlnMetArgProTyThrArgPheArgThrProGluPro..... 108
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
19708  CTTAGTTGAGGTACAGCTCGCGCATCTTATAGGGGACCCCTCACC 19757

109  AspaHisTyrraspPheCysLeuLeuPro 118
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
19758  TGATCATCATGATGAGTTTGCCTTATGCCT 19788

seq_name: gb_sts:G35294

seq_documentation_block: 477 bp DNA STS 02-OCT-1997
LOCUS G35294 human STS SHGC-37409, sequence tagged site.
DEFINITION G35294
ACCESSION G35294
VERSION G35294.1 GI:2459462
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 477)
AUTHORS Myers,R.M.
JOURNAL Unpublished (1997)
COMMENT
Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 415/7259687
Fax: 415/7259689
Email: myers@shgc.stanford.edu

Primer A: AACATCTTTCCATGAAAGTTGATG
Primer B: CTTTGGCATCTTCTTGCAA
STS size: 106
PCR Profile:
Initial incubation: 95 degrees C for 10 minutes
Denaturation: 94 degrees C for 15 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 23 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9600

Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
Amplitaq Gold Polymerase: 0.07 units/ul
Total Vol: 5 ul

Buffer: MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3

Prepared with primer pairs provided by Sandoz, derived from N51315
-- Washington University/Merck EST sequence.
FEATURES
source
1..477
/organism="Homo sapiens"
/db_xref="taxon:9606"
44..149
primer_bind
44..67

```

```

primer_bind 153 a 115 c 79 g 130 t
BASE COUNT
ORIGIN

alignment_scores:
Quality: 118.50 Length: 83
Ratio: 2.370 Gaps: 2
Percent Similarity: 60.241 Percent Identity: 33.735

alignment_block:
US-09-327-750D-34 x G35294/rev ..
Align seg 1/1 to reverse of: G35294 from: 1 to: 477

43  LysLysProGlyGlyAsnValArgLysValArgLysValArgLysValProAs 59
:::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||
474  CAGGAGCCTGGAGGAATGTTAAAGGGGTTTGGGCTCCACCTGCCCGGG 425

59  nPheLeuTrpAlaIleProAsnArgHisValAspHisSerGlu..... 73
:::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||
424  TTTTGGAGGAGTGTGCCCATAGCTTGTGATAACATTGATATGATAG 375

74  .....GlyGlyGluGluValGlyArgPheValGlyGlnValMetGluAla 88
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
374  ATGGAGATGGAGATGATATGACGGTTTCATGGAGAGAGATGAGAGAGCTA 325

89  LysArgHisSerLysGluGlnMetArgProTyThrArgPheArgTh 105
:::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||
324  AGAGGAAATAGGCACTTCAGTTGAGGTACAGCTGCGCATCTTAT 275

105  rProGluPro.....AspaHisTyrraspPheCysLeuLeuPro 118
:::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||
274  AGGGGACCTCTCCACCATGATCATGATGAGTTTGCCTTATGCCT 226

seq_name: gb_to:AF097439

seq_documentation_block: 785 bp mRNA ROD 13-APR-1999
LOCUS AF097439 Mus musculus brain expressed x-linked protein 2 (Bex2) mRNA,
DEFINITION complete cds.
ACCESSION AF097439
VERSION AF097439.1 GI:4580591
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 785)
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Bex1, a gene with increased expression in parthenogenetic embryos,
is a member of a novel gene family on the mouse X chromosome
JOURNAL Hum. Mol. Genet. 8 (4), 611-619 (1999)
MEDLINE 99172070
REMARK Erratum: [[published erratum appears in Hum Mol Genet 1999
May;8(5):943]]
REFERENCE 2 (bases 1 to 785)
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1998) Cancer Unit, Queensland Institute of
Medical Research, Herston Rd, Brisbane, QLD 4029, Australia
FEATURES
source
1..785
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="dbEST:W48832"
/chromosome="X"
/map="near Plp"
/dev_stage="embryo; 15.5 dpc"
1..785
/gene="Bex2"
139..528
gene
CDS

```



```
VERSION AF220189.1 GI:7689028
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 828)
JOURNAL Xiao.H., Song.H., Gao.G., Ren.S., Chen.Z. and Han.Z.
AUTHORS A novel gene expressed in human hypothalamus
REFERENCE 2 (bases 1 to 828)
JOURNAL Xiao.H., Song.H., Gao.G., Ren.S., Chen.Z. and Han.Z.
AUTHORS Direct Submission
TITLE Submitted (30-DEC-1999) Chinese National Human Genome Center at
JOURNAL Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-tech Park, Pudong,
Shanghai 201203, P. R. China
FEATURES
source Location/Qualifiers
1..828
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="hypothalamus"
167..544
/codon_start=1
/product="uncharacterized hypothalamus protein HBEX2"
/db_xref="GI:7689029"
/translation="MESKEKRVNSLWENAOENKEQVANKGEPLALPLDAGEYC
VPRGNRRFRVRQPILOYRDMHRLGEPQARMERMERIGEVEVRLMEKLRKQLS
HSLRAVSTDPHHDHDFCLMP"
BASE COUNT 256 a 170 c 221 g 181 t
ORIGIN

alignment_scores:
Quality: 112.50 Length: 128
Ratio: 1.424 Gaps: 5
Percent Similarity: 61.719 Percent Identity: 28.906

alignment_block:
US-09-327-750D-34 x AF220189
..
Align seg 1/1 to: AF220189 from: 1 to: 828

1 MetAlaSerLysValLysGlnValIleLeuAspLeuThrValGluLys... 16
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 ATGGAGTCCAAAGAGAACACGAGCAGTAACAGCTCAGCATGGAATGC 216
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 AspLysLysAsnLysLysGlyLysAlaSerLysGlnSerGluGluG 33
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
217 CAACCAAGAAATGAGAAAGGAGCAAGTTGCTAATAAGGGAGCCCT 266
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
33 LuSerHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnVal 49
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
267 TGGCCCTCCCTTTGGATGCTGGTGATACTGTGCTAGAGGAATCGT 316
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
50 ArgArg...LysValArgLeuValProAsnPheLeuTyrPalaIlePr 65
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
317 AGCGCGTTCCGCGTTAGCGCCATCTCTGCAGTATAGATGGATATGAT 366
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
65 oAsnArgHisValAsp.....HisSerGluGlyGlyGluG 77
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
367 GCATAGCTTGGAGACCAAGCAGGATGAGAGAGAGATATGGAAA 416
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
77 LuValGlyArgPheValGlyGlnValMetGluAlaLysArgHisSerLys 93
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
417 GGATTGGGGAGAGGTGAGACAGCTGATGGAAGAGCTGAGG..... 457
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
94 GluGlnGlnMetArgProTyrThrArgPheArgThrProGluPro..... 108
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
458 GAAAGACAGTTAGTCTAGTCTGCGGGGAGTCAGCAGTACCTGACCCCTCA 507
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
109 ....AspAsnHisTyrAspPheCysLeuIlePro 118
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

508 CCATGACCATCATGATGAGTTTGGCTTATGCC 541

FEATURES
source
Location/Qualifiers
1. 458
/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/clone_image="72056"
/clone_lib="Barstead MPLR81"
/sex="mixed"
/tissue_type="kidney"
/dev_stage="P0.5"
/note="Vector: p773D-Pac (Pharmacia) with a modified polylinker. Site 1: EcoRI; site 2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', 3']; double-stranded cDNA was ligated to Eco RI adaptors [CATGATTCGGTACC], digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library constructed by Bob Barstead."

BASE COUNT 147 a 88 c 115 g 108 t

ORIGIN

alignment_scores
Quality: 534.00 Length: 116
Ratio: 4.944 Gaps: 0
Percent Similarity: 93.103 Percent Identity: 86.207

alignment_block
US-09-327-750D-34 x AA473525

Align seg 1/1 to: AA473525 from: 1 to: 458

3 SerLysValLysGlnValIleuAspLeuThrValGluLysAspLys 19
13 GCGAAATTTAAACAGCTACTGCTCTGCTGCGAAGACAAA 62
19 aAnLysValGlyLysAlaSerLysGlnSerGluGluSerHis 36
63 AGACAAAAGAGTGGGAGGCTCCAAACAAAGTGAAGAACCCACC 112
36 lAluGluValGluAsnLysLysProGlyLysValArgLys 52
113 ATCTGGAGGCTGGAAACAGAGCTGGGGAAATGTCGAGGAA 162
53 ValArgLeuValProAsnPheLeuTrpAlaIleProAsnArgHis 69
163 GTCAGCGGACTTGTGCTTAACCTTCTGCGCCATACCAATAGCATGT 212
69 lAspHisSerGluGlyGluValGlnValGlnValGlnVal 86
213 TGATCGCATGAGGGGAGGAGGATTTGGAGATTGTAGTCAGGAA 262
86 etGluAlaValArgHisSerLysGluGlnSerGlyProThrArg 102
263 CAGAGTCAAGAGAAAGACTACGAGCAGCAGGTAGCGCTACAGCGCT 312
103 PheArgThrProGluProAsnHisLysArgPheCysLeuIlePro 118
313 TTCCGAACCCCGAACCTGACAATATTACGACTTTTGCCTCATCT 360

seq_name: gb_est1:AW251884

seq_documentation_block:
LOCUS AW251884.1 540 bp mRNA EST 17-DEC-1999
DEFINITION UT-R-B30-adn-b-06-0-UI-a UT-R-B30 Rattus norvegicus cDNA clone
UT-R-B30-adn-b-06-0-UI-3', mRNA sequence.
ACCESSION AW251884
VERSION AW251884.1 GI:6595475
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
1 (bases 1 to 540)
AUTHORS
Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE
Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
MEDLINE
97044477
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
Department of Molecular Biology
451 East 17th Street
Madison Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.wisc.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized atrium at 16.5 dpc library cDNA library Preparation: 16.5 dpc, atrium. One distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA-Yes.

FEATURES
Location/Qualifiers
1. 540
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UT-R-B30-adn-b-06-0-UI"
/clone_lib="UI-R-B30"
/dev_stage="adult"
/adaptor="BIOB (Life Technologies)"
/polylinker="Site 1: Not I; Site 2: Eco RI; The UI-R-B30 library is a subtracted library derived from the UI-R-AAL, UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AS1, UI-R-AF1, and UI-R-AG1 libraries. These libraries represent tissues from rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV canal at 15 dpc, and ventricle at 13 dpc. The tag is a string of 5-6 nucleotides present at the Not I site and the oligo-dT track. The library was constructed as a subtracted library. The library was constructed as 6-791-806, 1996.
TAG_L1B=UI-R-B30.
TAG_TISSUE=atrium at 16.5 dpc
TAG_SEQ=GATTC

BASE COUNT 146 a 120 c 94 g 180 t

ORIGIN

alignment_scores
Quality: 520.00 Length: 97
Ratio: 5.417 Gaps: 0
Percent Similarity: 98.969 Percent Identity: 98.969

alignment_block
US-09-327-750D-34 x AW251884/rev

Align seg 1/1 to reverse of: AW251884 from: 1 to: 540

22 LysGlyLysAlaSerLysGlnSerGluGluSerHisLeucl 38
538 AAGGTGGGAGGCTCCAAACAAAGTGAAGAGGATCCCATCTGGA 489
38 uGluValGluAsnLysLysProGlyLysValArgLysValArg 55
488 AGAGTTTGAACAAAGAGCTGGGGCAATGTCAGGAGGAAGTCAGGC 439
55 rGluValProAsnPheLeuTrpAlaIleProAsnArgHisValAspHis 71

/clone="IMAGE:538606"
 /clone_lib="Beddington mouse embryonic region"
 /sex="pooled"
 /tissue_type="embryo"
 /dev_stage="7.5dpc"
 /lab_host="BRL25"
 /note="gen: whole embryo; Vector: pCMV-SPORT; Site_1: 5' end; Site_2: 3' end; Cloning strategy: PCR; Oligo dT: Gastrulating embryos were collected at 7.5dpc from C57BL/6 x DBA matings, excluding embryos that had developed head folds and all extraembryonic tissues. Average insert size: 1.3 kb (range: 0.5 - 3.0 kb). Referenced in Development 121, 2479-2489 (1995)"

BASE COUNT 110 a 56 c 92 g 54 t
 ORIGIN

alignment_scores: Quality: 444.00 Length: 100
 Ratio: 4.770 Gaps: 0
 Percent Similarity: 93.000 Percent Identity: 86.000

alignment_block:

US-09-327-750D-34 x AAL17429 ..

Align seg 1/1 to: AAL17429 from: 1 to: 312

1 MetAserLysValLysGlnValLleuAspLeuThrValGluLysAs 17
 |||||
 12 ATGGATCCAAATTATACAGTCATCTACTGTGGAGAAAGA 61
 17 PylsAsnLysGlyGlyLysAlaSerLysGlnSerGluGluGlu 34
 |||||
 62 CAAAGAGCAAGAGGTGGAGGCTCCAGAGAGAGAGAGAGAC 111
 34 erHisLysGluGluValGluGlnLysValProGlyGlnValArg 50
 |||||
 112 CCACCATCTGGAGAGGTTGAAACAGAGAGCTGGGGGAAATGTCCGA 161
 51 ArgLysValArgLeuValProAsnPheLeuTrpAlaLeuProAsnAr 67
 |||||
 162 AGGAAGTCAGGCACCTGTGGCTAACTTCTCTGGGCATACCTAATAG 211
 67 gHisValAspHisSerGluGlyGlyGluValGlyArgPheValGly 84
 |||||
 212 GCATGTGATCCATCAGAGGAGAGAGGAGGAGGAGGAGGAGGAG 261
 84 InValMetGluAlaLysArgHisSerLysGluGlnGlnMetArgProTyr 100
 |||||
 262 AGGACAGAGTCAAGAGAGAGTACGAGGAGCAGCAGGTGAGGCCTTAC 311

seq_name: gb_est1:A1012535

seq_documentation_block: 444 bp mRNA EST 15-JUN-1998
 LOCUS A1012535
 DEFINITION EST206986 Normalized rat placenta, Bento Soares Rattus sp. cDNA

ACCESSION A1012535

VERSION A1012535

KEYWORDS EST

SOURCE Rattus sp.

ORGANISM Rattus sp.

REFERENCE 1 (bases 1 to 444)

AUTHORS Lee N.H., Glodek A., Chandra, I., Mason, T.M., Quackenbush, J.,

TITLE Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat

JOURNAL Unpublished (1998)

CONTACT: Lee, NH

COMMENT The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13-21.

FEATURES

source

Location/Qualifiers
 ..444..Battus sp.
 /clone="IMAGE:538606"
 /clone_lib="Beddington mouse embryonic region"
 /sex="pooled"
 /tissue_type="embryo"
 /dev_stage="7.5dpc"
 /lab_host="BRL25"
 /note="gen: whole embryo; Vector: pCMV-SPORT; Site_1: 5' end; Site_2: 3' end; Cloning strategy: PCR; Oligo dT: Gastrulating embryos were collected at 7.5dpc from C57BL/6 x DBA matings, excluding embryos that had developed head folds and all extraembryonic tissues. Average insert size: 1.3 kb (range: 0.5 - 3.0 kb). Referenced in Development 121, 2479-2489 (1995)"

BASE COUNT 135 a 104 c 81 g 124 t
 ORIGIN

alignment_scores: Quality: 426.00 Length: 80
 Ratio: 5.462 Gaps: 0
 Percent Similarity: 97.500 Percent Identity: 97.500

alignment_block:

US-09-327-750D-34 x A1012535/rev ..

Align seg 1/1 to reverse of: A1012535 from: 1 to: 444

39 GluValGluAsnLysLysPheGlyGlyAsnValArgLysValArgAr 55
 |||||
 444 GAGGTGAAACATGAAGCCTGGGGGCAATGTTCAGGAGGAAAGTCAGGCG 395
 55 GluValProAsnPheLeuTrpAlaLeuProAsnArgHisValAspHis 72
 |||||
 394 ACTGTGCTTACTTCTTGTGGGCTTCTTGTAGGCTTGTAGGCTTGTAGGCT 345
 72 erGluGlyGluGluValGlyArgPheValGlyGluValMetGluAla 88
 |||||
 344 GTGAAGGGGAGAGAGGTTGGGAGATTCGTAGGCGAGGTGATGGAAGCC 295
 89 LysArgHisSerLysGluGlnGlnMetArgProTyrThrArgPheArgH 105
 |||||
 294 AAGAGAAAGTCTTAAGGAGCAACAGATGAGGCTTTACAGCGGTTTCGGAAC 245
 105 rProGluProAsnHisTyrAspPheCysLeuLeuPro 118
 |||||
 244 CCTGACCTGACATCATCATTACGCTTTGGCTCATCTACT 205

seq_name: gb_est2:M35893

seq_documentation_block: 472 bp mRNA EST 14-MAY-1996

LOCUS M35893
 DEFINITION mc5909.r1 Soares mouse embryo NM013.5 14.5 Mus musculus cDNA
 clone IMAGE:352280 5' similar to SW:HC74_HUMAN Q00994 OVARIAN
 GRANULOSA CELL 13.0 KD PROTEIN HG74. [1] ; mRNA sequence.

ACCESSION M35893

VERSION M35893

KEYWORDS EST

SOURCE Mouse

ORGANISM Mouse

REFERENCE 1 (bases 1 to 472)

AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

Waterston, R.

TITLE The WashU-RHMI Mouse EST Project

JOURNAL Unpublished (1996)

CONTACT: Marra M/Mouse EST Project

COMMENT Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

```
34 erHlHlaLeuGlulValGluAnLysPtoGlyGlyAenValArg 50
|||||
268 CCCACCATCTGGAAGAGGTGAAACAGAGAGCCCTGGGGGAAATGTCGGA 317
|||||
51 ArgLysValArgArgLeuValProAsnPhelLeuTrpAlaIleProAsnAr 67
|||||
318 AGAAGTC-AGGCACCTGTGCTTAACCTTCTCTGGGCCATACCAATAG 366
|||||
67 gHlValAspHisSerGluGlyGluGluValGlyArgPheValGlyG 84
|||||
367 GCATGTTCATCCATGAGGAGGAGAGGTGTGGAGATTGTGTGC 416
|||||
84 InValMetGluAlaLysArgHisSerGluGluGluGluMetArgPro 99
|||||
417 AGGAACAGAGTCAAGAGAGTACGAGCAGCAGGTGAGGCT 463
|||||
seq_name: gb_est1:AL118340

seq_documentation_block:
LOCUS AL118340 422 bp mRNA EST 23-SEP-1999
DEFINITION V9112026 Beddington mouse dissected endoderm Mus musculus cDNA
clone 528_13N20 5', mRNA sequence.
ACCESSION U011834
VERSION AL118340.1 GI:5920179
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
Harrison.S.M., Dunwoodie, S.L., Arkell, R.M., Lehrach, H. and
Beddington, R.S.
Isolation of novel tissue-specific genes from cDNA libraries
representing the individual tissue constituents of the gastrulating
mouse embryo.
JOURNAL Development 121 (8), 2479-2489 (1995)
MEDLINE 95401865
COMMENT Contact: Ruiz, P., Lehrach, H. and Avner, P.
ECF Mouse Transcript Mapping Consortium
Genoscope - CNS
2, rue Gaston Cremieux, CP 5706, 91057 Evry CEDEX, France
Email: pavner@pasteur.fr
Clone available from Ressourcenzentrum, Heubnerweg 6, D-14059
Berlin, Germany. Web site http://www.rzpd.de
Seq primer: CCGGCGGAGATCCCGGT.
Location/Qualifiers
location="Mus musculus"
organism="Mus musculus"
strain="C57Bl6 x DBA"
db_xref="taxon:10090"
clone="528_13N20"
tissue_type="dissected endoderm"
dev_stage="7.5 dpc"
note="Vector: pSport1; Site_1: NotI; Site_2: SalI; Cloned
unidirectionally. - High quality sequence only submitted. -
Average insert size: 1.2 kb (range: 0.2 - 2 kb)."
BASE COUNT 128 a 86 c 143 g 85 t
ORIGIN

alignment_scores:
Quality: 407.00 Length: 90
Ratio: 4.788 Gaps: 0
Percent similarity: 94.444 Percent identity: 87.778

alignment_block:
US-09-327-750d-34 x AL118340 ..
Align seg 1/1 to: AL118340 from: 1 to: 422
1 MetAlaSerLysValLysGlnValIleLeuAspLeuThrValGluLysAs 17
```

```
|||||
146 ATGGCATCCCAATTTAAACAGTCAATCGATCTCCTGCGGAGAGAG 195
|||||
17 pLyLysAnLysLysGlyGlyLysPAlaSerLysGlnSerGluGluG 34
|||||
196 CAAAAAGACAAAAGAGAGGTGGGAGGCTCCAAAACAAAGTGAAGAGA 245
|||||
34 erHlHlHlaLeuGluGluValGluAnLysPtoGlyGlyAenValArg 50
|||||
246 CCCACCATCTGGAAGAGGTGAAACAGAGAGCCCTGGGGGAAATGTCGGA 295
|||||
51 ArgLysValArgArgLeuValProAsnPhelLeuTrpAlaIleProAsnAr 67
|||||
296 AGGAATCTCAGCGACCTGTGCTTAACCTTCTCTGGGCCATACCAATAG 345
|||||
67 gHlValAspHisSerGluGlyGluGluValGlyArgPheValGlyG 84
|||||
346 GCATGTTCATCCCAATCGAGGAGAGGATGTTGGGAGATTGTGTGTCGC 395
|||||
84 InValMetGluAlaLysArg 90
||
396 AGGAACAGAGTCAAGAGA 415
```

508 AGCGTTCCGACCCGGAACCTGACATCATACGACTTTGCTCAT 557
117 epro 118
558 ACCT 543
seq_name: gb_est2:BG083261
seq_documentation_block: 792 bp mRNA EST 26-JAN-2001
LOCUS BG083261 H3086C08-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
DEFINITION H3086C08 5', mRNA sequence.
ACCESSION BG083261
VERSION BG083261.1 GI:12565829
KEYWORDS EST.
ORGANISM mouse mouse.
MUSCULUS: Mus musculus.
Eukaryote: Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 792)
Kargul G.J., Dudekula D.B., Qian Y., Lim M.K., Jaradat S.A., Tanaka
, T.S., Carter M.G. and Ko, M.S.H.
Verification and Initial Annotation of NIA mouse 15K cDNA clone set
Unpublished (2001)
Other ESTs: H3086C08-3
Contact: George J. Kargul
Laboratory of Genetics
National Institute of Health
333 National Institute Suite 400
Email: cdna@nigms.nih.gov Baltimore, MD 21224-6820, USA
This clone set has been freely distributed to the community. Please
visit <http://lgaun.grc.nia.nih.gov/cDNA/15k.html> for details.
Plate: H3086 row: C column: 08
Seq primer: -21M13 Reverse
High quality sequence stop: 792
POLYA-No.
Location/Qualifiers
1. 792
/organism="Mus musculus"
/strain="C57BL/6J"
/db="EST"
/clone="H3086C08"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_stage="Clones arrayed from a variety of cDNA libraries"
/lab_host="DH10B"
/note="Vector: pSPORT1. Site 1: SalI; site 2: NotI; This
clone is among a rearranged set of 15,247 clones from 11
embryo cDNA libraries (including preimplantation stage
embryos from unfertilized egg to blastocyst, embryonic
part of 81.5 embryos, extraembryonic part of 27.5 embryos
and 81.5 female mouse embryos). Size 15 kb. All
source cDNA libraries are cloned unidirectionally with Oligo(dT
)-Not primers. References include: (1) Genome-wide
expression profiling of mid-gestation placenta and embryo
using a 15,000 mouse developmental cDNA microarray, 2000,
Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)
Large-scale cDNA analysis reveals phased gene expression
patterns during preimplantation mouse development, 2000,
Development, 127: 1737-1749; (3) Genome-wide mapping of
unselected transcripts from extraembryonic tissue of
7.5-day mouse embryos reveals enrichment in the t-complex
and under-representation on the X chromosome, 1998, Hum
Mol Genet 7: 1707-1718.
BASE COUNT 247 a 153 c 210 g 179 t 3 others
ORIGIN
alignment_scores:
Quality: 546.00 Length: 118
Ratio: 4.964 Gaps: 0

Percent Similarity: 93.220 Percent Identity: 87.288
alignment_block:
US-09-327-750d-34 x BG083261
Align seg 1/1 to: BG083261 from: 1 to: 792
1 MetAlaSerLysValLysGlnValLysLeuAspLeuThrValGluLysAs 17
|||||
190 ATGGCATCCAAATTAACAAAGTCACTGATCTCACTGATGGAGAAGA 239
17 PLYsLysAsnLYsGlyGlyGlyAlaSerLysGlnSerGluGluGus 34
|||||
240 CAAAAGACAAAAGAGGGGGGAGGCTCCAAACAAAGTGAAGAGAAC 289
34 eRfHhHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 50
|||||
290 CCCACATCTGGAGAGAGTGGAAAGAGAGCTGGGGGAAATGTCGA 339
51 ArgLYsValArgGlyGlyValProAsnPheLeuArgPAlaLeProAsnAr 67
|||||
340 AGGAAGTCAAGCGACTTCTGCTTAACCTTCTCTGGGCCATACCAATAG 389
67 gHisValAspHisSerGluGlyGlyGluGluValGlyArgPheValGlyG 84
|||||
390 GCATGTGTATCCCAATGAAGAGGGAGAGGATGTTGGAGATTGTAGTGC 439
84 LnvAlMetGluLysArgHisSerLysGluGluGlnMetArgProTyr 100
|||||
440 AGGGAACAGAGTCCAGAGAAAGTACAGGAGACAGCCAGTGGCGCTTAC 489
101 ThrArgPheArgThrProGluProAsnHisLysArgPheCysLeuL 117
|||||
490 AGGCGTTTCGGAACCCGGAACCTGACATCATACGACTTTGCTCAT 539
117 epro 118
540 ACCT 543
seq_name: gb_est1:AA473525
seq_documentation_block: 458 bp mRNA EST 18-JUN-1997
LOCUS AA473525 H3086C08-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
DEFINITION v978a09.t1 Barstead MPRB1 Mus musculus cDNA clone IMAGE:174056 5',
similar to SM:HG74_HUMAN Q00994 OVARIAN GRANULOSA CELL 13.0 KD
PROTEIN HGR74. ?, mRNA sequence.
ACCESSION AA473525
VERSION AA473525.1 GI:2201752
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 458)
Marlet S., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque, T.,
Giesel S., Kucaba T., Lacy M., Le M., Martin J., Morris M.,
Schellenberg K., Steptoe M., Tan F., Underwood K., Moore B.,
Theising B., Wylie T., Lennon G., Soares B., Wilson R. and
Waterston R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Maria N/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1801
Email: mouse@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:511536
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 417.

```

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus 1 to 540)
Bonaldo,M.F., Lennon,G. and Soares,M.B
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
Department of Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 9250
Fax: 319 335 9565
Email: mscoates@blue.weeg.uiowa.edu
The sequence of an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
Normalized atrium at 16.5 dpc library. cDNA Library preparation:
Mammalian oocyte injection and translation. Details will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.
Location/Qualifiers
1..540
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/cclone="UI-R-BJ0-adn-b-06-0-UI"
/cloned_into="UI-R-BJ0"
/clone_size="118"
/lab_hosts="PHI08 (Life Technologies)"
/notes="vector: pT730-Pac (Pharmacia) with a modified
polylinker: Site 1: Not I; Site 2: Eco RI; The UI-R-BJ0
library is a subtracted library derived from the UI-R-AAL,
UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AFL, and
UI-R-AG1 libraries. These libraries represent tissues from
rat atrium at 16.5 dpc, ventricle at 15 dpc, AV canal
at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
canal at 15 dpc, and the heart of the embryo. The
string of 5-6 nucleotides present between the NotI site
and the oligo-dT track. The library was constructed as
described by Bonaldo, Lennon and Soares, Genome Research
6: 791-806, 1996.
TAG_L1B=UI-R-BJ0
TAG_TISSUE=atrium at 16.5 dpc
TAG_SEQ=GATTC 94 g 180 t
..a..120 c
BASE COUNT
ORIGIN

alignment_scores
Quality: 520.00 Length: 97
Ratio: 5.417 Gaps: 0
Percent Similarity: 98.969 Percent Identity: 98.969

alignment_block
US-09327-750D-34 x AW251884/rev ..
Align seq 1/1 to reverse of: AW251884 from: 1 to: 540

22 LysGlyGlyLysAlaSerLysGlnSerGluGluGluSerHisHisLeuGI 38
538 AAAGGTGGGAAGGCCCTCCAAACAAAGTAGAGGAATCCACCACATCTGGA 489
38 uGUuAUGuAnLysLysAprGtGtGlyAenValArgArgLysValArgA 55
488 AGGGTGTGAACAGAACAGACCCCTGGGGGGCGATGTCAGAGGAATCCAGC 439
55 gGUuValProAenPhuLzuvTPhaLzLeProAenArgHisValAeopHis 71

```


OM of: US-09-327-750d-35 to: EST.* out_format : pfs

Date: Mar 11, 2002 2:16 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framet_p2n.model -DEV=xlh
-O/cgnt2.1/USPTO.spool/US09327750/runtat_11032002_101153_20308/app_query.fasta_1.1472
-DB=EST -OFT=fastp -SURF=1rst -CAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOPEL=0.000 -LOPEXT=0.000 -CAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-FQPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELETE=7.000 -START=1 -MATRIX=blotsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09327750 -RCGNI_1.5654
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-327-750d-35

Query length: 117

Database: EST.*

Database sequences: 11351937

Database length: 1077921985

Search time (sec): 4085.940000

score_list:

Sequence	Strd	Orig	zScore	Escore	Len	Documentation
gb_hlc:AK010400	+	611.50	1049.04	2.9e-49	789	AK010400 Mus musculus ES cells
gb_est2:BG083261	+	611.50	1049.02	2.9e-49	792	AK083261 H3086C08-5 NIA Mouse
gb_est1:AA473525	+	605.00	1041.70	7.4e-49	458	AA473525 V978a09.r1 Barstead M
gb_est2:BG070341	+	575.50	987.47	7.7e-46	747	I BG070341 H3086C08-3 NIA Mouse
gb_est2:BF723075	+	568.50	977.99	2.6e-45	519	BF723075 mab27el2.y1 Soares_NME
gb_est1:BE333774	+	544.50	937.49	4.7e-43	463	BE333774 us27h06.y1 Soares_NME
gb_est2:W17712	+	541.00	934.14	7.2e-43	317	W17712 mb77e04.r1 Soares mouse
gb_est1:A1227867	+	539.50	927.48	1.7e-42	568	I A1227867 EST224562 Normalized
gb_est2:BI289546	+	511.50	878.38	9.1e-42	643	BI289546 UI-R-DKO-cff-c-12-0-U
gb_est1:A117429	+	503.50	869.71	2.8e-39	312	AA117429 mn21e12.r1 Beddington
gb_est2:W35893	+	484.50	834.09	2.7e-37	472	W35893 mc3g09.r1 Soares mouse
gb_est2:BF722027	+	479.00	828.16	5.8e-37	286	BF722027 mab27el2.x1 Soares_NME
gb_est2:W54487	+	469.50	808.36	7.3e-36	465	W54487 md09cll.r1 Soares mouse
gb_est1:AL118340	+	463.50	798.74	2.5e-35	422	AL118340 V9112b26 Beddington m
gb_est1:AW251884	+	459.00	789.25	8.5e-35	540	AW251884 UI-R-BJO-adn-b-06-0-U
gb_est1:AA104150	+	426.50	734.87	9.1e-32	433	AA104150 mc50h03.r1 Life Tech
gb_est2:W85403	+	426.50	734.73	9.2e-32	442	W85403 mf58d02.r1 Soares mouse
gb_est1:A1012535	+	373.00	642.62	1.2e-26	444	I A1012535 EST206986 Normalized
gb_est1:AV137854	+	368.00	636.78	2.6e-26	300	AV137854 AV137854 Mus musculus
gb_est1:AA516739	+	365.00	631.85	5.0e-26	230	AA516739 v84c04.r1 Knowles Sol
gb_est2:W61422	+	363.50	630.70	5.8e-26	237	W61422 md29g02.r1 Soares mouse
gb_est1:A1517301	+	335.50	609.93	8.3e-25	639	I A1517301 AL517301 LTI_NFL011_NH
gb_est2:BE883414	+	335.50	609.07	9.2e-25	722	BE883414 601509167F1 NIH_MGC_71
gb_est1:AL544124	+	335.50	607.95	1.1e-24	845	I AL544124 AL544124 LTI_NFL006.PI
gb_est1:A117899	+	335.50	607.79	1.1e-24	865	I A117899 AU17899 HEMBA1 Homo
gb_est2:BG122197	+	335.50	607.78	1.1e-24	866	BG122197 602349748F1 NIH_MGC_90
gb_est1:AL545120	+	335.50	607.47	1.1e-24	905	I AL545120 AL545120 LTI_NFL006.PI
gb_est2:BG709427	+	335.50	607.29	1.2e-24	929	BG709427 602673481F1 NIH_MGC_96
gb_est2:BG035675	+	335.50	606.94	1.2e-24	975	BG035675 602325745F1 NIH_MGC_90
gb_est1:AL560850	+	335.50	606.89	1.2e-24	983	I AL560850 AL560850 LTI_NFL010_BC
gb_est1:AL539004	+	335.50	606.61	1.3e-24	1023	I AL539004 AL539004 LTI_NFL010_FH
gb_est1:AL549875	+	335.50	606.46	1.3e-24	1044	I AL549875 AL549875 LTI_NFL006.FH
gb_est2:BF930215	+	331.00	602.88	2.0e-24	579	BF930215 IL5-N02027-111200-319
gb_est1:AL534630	+	330.50	602.18	2.2e-24	566	I AL534630 AL534630 LTI_FLO13_FH
gb_est1:AL523320	+	349.00	595.28	5.4e-24	1043	I AL523320 AL523320 LTI_NFL003_N
gb_est1:AW385267	+	347.50	597.43	4.1e-24	534	I AW385267 RC025727F1 NIH_MGC_59
gb_est2:BG527431	+	347.50	594.82	5.7e-24	772	BG527431 60255727F1 NIH_MGC_59
gb_est2:BG009563	+	346.00	595.11	5.5e-24	581	I BG009563 QV1-GN0319-021200-526
gb_est1:AL582060	+	346.00	590.16	1.0e-23	1037	I AL582060 AL582060 LTI_NFL010_FH
gb_est2:BI040887	+	344.50	592.88	7.4e-24	489	I BI040887 QV3-NT0278-120201-502

gb_est1:AL573904 - 344.50 586.65 1.6e-23 1182 I AL573904 AL573904 LTI_NFL010
gb_est2:BI040889 - 340.50 584.98 2.0e-23 565 I BI040889 QV3-NT0278-120201-
gb_est1:AW934883 - 338.50 580.35 3.7e-23 668 I AW934883 RCI-DF0001-230200-
gb_est1:AW934905 - 338.50 580.18 3.8e-23 685 I AW934905 RCI-DF0001-230200-
gb_est1:AW934769 - 338.00 579.56 4.1e-23 662 I AW934769 RCI-DF0001-230200-
seq_name: gb_hlc:AK010400

seq_documentation_block:

LOCUS AK010400 789 bp mRNA HTC 05-JUL-2001
DEFINITION Mus musculus ES cells cDNA, RIKEN full-length enriched library,
clone:2410004M13, full insert sequence.
ACCESSION AK010400
VERSION AK010400.1 GI:12845816
KEYWORDS CAP Trapper.
SOURCE Mus musculus
clone:2410004M13
clone:2410004M13
clone:2410004M13

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 789)
Carninci.P. and Hayashizaki.Y.

REFERENCE

AUTHORS Carninci.P. and Hayashizaki.Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PubMed 10349636

REFERENCE

2 (bases 1 to 789)
Carninci.P., Shibata.Y., Hayatsu.N., Sugahara.Y., Shibata.K.,
Itoh.M., Konno.H., Okazaki.Y., Muramatsu.M. and Hayashizaki.Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome research. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PubMed 11042159

REFERENCE

3 (bases 1 to 789)
Shibata.K., Itoh.M., Aizawa.K., Nagaoka.S., Sasaki.N., Carninci.P.,
Konno.H., Akiyama.J., Nishi.K., Kitsuai.T., Tashiro.H., Itoh.M.,
Sumi.N., Ishii.Y., Nakamura.S., Hazama.M., Nishine.T., Harada.A.,
Yamamoto.R., Matsumoto.H., Sakaguchi.S., Ikegami.T., Kashiwagi.K.,
Fujiwara.S., Inoue.K., Togawa.Y., Izawa.M., Ohara.E., Watahiki.M.,
Yoneda.Y., Ishikawa.T., Ozawa.K., Tanaka.T., Matsura.S., Kawai.J.,
Okazaki.Y., Muramatsu.M., Inoue.Y., Kira.A. and Hayashizaki.Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384-multipipillary sequencer
JOURNAL Genome research. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PubMed 11076861

REFERENCE

4 (bases 1 to 789)
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 789)

AUTHORS

Adachi.J., Aizawa.K., Akahira.S., Akimura.T., Aono.H., Arai.A.,
Arakawa.T., Carninci.P., Fukuda.S., Fukunishi.Y., Furuno.M.,
Hanagaki.T., Hara.A., Hayatsu.N., Hiramoto.K., Hiraoka.T., Horii.F.,
Imotani.K., Ishii.Y., Itoh.M., Izawa.M., Kato.H., Kawai.J.,
Kojima.Y., Konno.H., Kouda.M., Koya.S., Kurihara.C., Matsuyama.T.,
Miyazaki.A., Nishi.K., Nomura.K., Numazaki.R., Ohno.M., Okazaki.Y.,
Okido.T., Owa.C., Saito.H., Saito.R., Sakai.C., Sakai.K., Sano.H.,
Sasaki.D., Shibata.K., Shibata.Y., Shinagawa.A., Shiraki.T.,
Sogabe.Y., Suzuki.H., Tagami.M., Tagawa.A., Takahashi.F.,
Tanaka.T., Tejima.Y., Toya.T., Yamamura.T., Yasunishi.A.,
Yoshida.K., Yoshino.M., Muramatsu.M. and Hayashizaki.Y.

TITLE

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome.res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/ Tel:81-45-503-9222,
Fax:81-45-503-9216)

COMMENT

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAGAGATCCCAAGACTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAGAGATCTCGAGTTAATAATTAATCCCGCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.

Host: SOLR.

FEATURES

source

Location/Qualifiers

1..789
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /db_xref="MGD:MGI:1328321"
 /db_xref="MGD:MGI:1908964"
 /clone="241004M13"
 /cell_type="ES cells"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 208..564
 /note="putative"
 /codon_start=1
 /protein_id="BAB26911.1"
 /db_xref="GI:12845817"
 /translation="MASKFQVILDTVEKDKKDKGKSKQSEEPHLEEVENKK
 PGGVRRKVRRLVPNFWLWAIPIRHRVDNREGDVGFRVVOGTVEVRKRTTEQOVRPYR
 FRTEPDNHYDECLIP"
 770..775
 /note="putative"
 789
 /note="putative"

CDS

208..564
 /note="putative"

polyA_signal

770..775
 /note="putative"

polyA_site

789
 /note="putative"

BASE COUNT

239 a 161 c 205 g 184 t

ORIGIN

1..789

alignment_scores:

Quality: 611.50 Length: 118
 Ratio: 5.226 Gaps: 1
 Percent Similarity: 99.153 Percent Identity: 98.305

alignment_block:

US-09-327-750D-35 x AK010400

Align seg 1/1 to: AK010400 from: 1 to: 789

1 MetalaSerLys...LysGlnValIleLeuAspLeuThrValGluLysAs 16
 208 ATGCATCCCAATTTAAACAGTCATCTGGATCTACTGTGGAGAAGA 257
 16 pLysLysAspLysArgGlyGlyLysAlaSerLysGlnSerGluGluGlu 33
 258 CAAAAGACAAAAAGGTGGGAAGCCCTCCCAACAAAGTGAAGAAGAAC 307
 33 roHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 49
 308 CCCACCATCTGGGAAGAGGTGAAACAAAGACCCCTGGGGGAAATCTCGA 357
 50 ArgLysValArgArgLeuValProAsnPheLeuTTPAlaIleProAsnAr 66
 358 AGGAAGTCAGGCGACTTGTGCTAACTTCTCTGGGCCATACCAATAG 407
 66 gHisValAspArgAsnGluGlyGlyGluAspValGlyArgPheValValG 83
 408 GCATGTTGATCGCAATGAAGGGGGAGAGATGTTGGGAGATTGTAGTGC 457
 83 lngLysThrGluValLysArgLysThrThrGluGlnGlnValArgProTyr 99

458 AGGGAACAGAGAGTCAAGAGAAAGACTACGGAGCAGAGGTGAGGCCTTAC 507

100 ArgArgPheArgThrProGluProAspAsnHisTyrAspPheCysLeuI 116

508 AGGGGTTCCGAACCCCGGAACCTGACAATCATTAGACTTTTGGCCTCAT 557

116 ePro 117

|||||

558 ACCT 561

seq_name: gb_est2:BC083261

seq_documentation_block:

LOCUS BC083261 792 bp mRNA EST 26-JAN-2001
 DEFINITION H3086C08-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
 H3086C08 5', mRNA sequence.

ACCESSION BC083261 GI:12565829

VERSION BC083261

KEYWORDS EST.

SOURCE Mus musculus.

ORGANISM Mus musculus.

REFERENCE 1 (bases 1 to 792)

AUTHORS Kargul G.J., Dudekula D.B., Qian Y., Lim M.K., Jaradat S.A., Tanaka

T.S., Carter M.G. and Ko M.S.H.

Verification and initial annotation of NIA mouse 15K cDNA clone set

Unpublished (2001)

Other ESTs: H3086C08-3

Contact: George J. Kargul

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@lgsun.grc.nia.nih.gov

This clone set has been freely distributed to the community. Please

visit <http://lgsun.grc.nia.nih.gov/cDNA/15K.html> for details.

Plate: H3086 row: C column: 08

Seq primer: -21M13 Reverse

High quality sequence stop: 792

POLYA-NO.

Location/Qualifiers

1..792

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="H3086C08"

/clone_lib="NIA Mouse 15K cDNA Clone Set"

/sex="Clones arrayed from a variety of cDNA libraries"

/dev_stage="Clones arrayed from a variety of cDNA libraries"

/lab_host="DH10B"

/note="Vector: pSPORT1; Site1: SalI; Site2: NotI; This

clone is among a rearranged set of 15,247 clones from 11

embryo cDNA libraries (including preimplantation stage

embryos from unfertilized egg to blastocyst, embryonic

part of E7.5 embryos, extraembryonic part of E7.5 embryos

, and E12.5 female mesonephros/gonad) and one newborn

ovary cDNA library. Average insert size 1.5 kb. All

source libraries are cloned unidirectionally with Oligo(dT

)-Not primers. References include: (1) Genome-wide

expression profiling of mid-gestation placenta and embryo

using a 15,000 mouse developmental cDNA microarray, 2000,

Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)

Large-scale cDNA analysis reveals phased gene expression

patterns during preimplantation mouse development, 2000,

Development, 127: 1737-1749; (3) Genome-wide mapping of

unselected transcripts from extraembryonic tissue of

7.5-day mouse embryos reveals enrichment in the t-complex

and under-representation on the X chromosome, 1998, Hum

Mol Genet 7: 1967-1978.

247 a 153 c 210 g 179 t 3 others

BASE COUNT

ORIGIN


```

VERSION      BG070341.1  GI:12552910
KEYWORDS     EST.
SOURCE       house mouse.
ORGANISM     Mus musculus

REFERENCE    Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE        Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A., Tanaka
JOURNAL      1 (bases 1 to 747)
COMMENT      T.S., Carter, M.G. and Ko, M.S.H.
              Verification and initial annotation of NIA mouse 15K cDNA clone set
              Unpublished (2001)
              Other_ESTs: H3086C08-5
              Contact: George J. Kargul
              Laboratory of Genetics
              National Institute on Aging/National Institutes of Health
              333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
              Email: cdna@lgsun.grc.nia.nih.gov
              This clone set has been freely distributed to the community. Please
              visit http://lgsun.grc.nia.nih.gov/cDNA/15K.html for details.
              Plate: H3086 row: C column: 08
              Seq primer: -21M13 forward
              High quality sequence stop: 747
              POLYA=Yes.

FEATURES     source
             1..747
                Location/Qualifiers
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="H3086C08"
                /sex="Clones arrayed from a variety of cDNA libraries"
                /dev_stage="Clones arrayed from a variety of cDNA
                libraries"
                /lab_host="PH10B"
                /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                clone is among a rearranged set of 15,247 clones from 11
                embryo cDNA libraries (including preimplantation stage
                embryos from unfertilized egg to blastocyst, embryonic
                part of E7.5 embryos, extraembryonic part of E7.5 embryos
                , and E12.5 female mesonephros/gonad) and one newborn
                ovary cDNA library. Average insert size 1.5 kb. All
                source libraries are cloned unidirectionally with Oligo(dT
                )-Not primers. References include: (1) Genome-wide
                expression profiling of mid-gestation placenta and embryo
                using a 15,000 mouse developmental cDNA microarray, 2000,
                Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)
                Large-scale cDNA analysis reveals phased gene expression
                patterns during preimplantation mouse development, 2000,
                Development, 127: 1737-1749; (3) Genome-wide mapping of
                unselected transcripts from extraembryonic tissue of
                7.5-day mouse embryos reveals enrichment in the t-complex
                and under-representation on the X chromosome, 1998, Hum
                Mol Genet 7: 1967-1978."
BASE COUNT   178 a 198 c 140 g 231 t
ORIGIN

alignment_scores:
  Quality: 575.50      Length: 118
  Ratio: 5.004         Gaps: 1
  Percent Similarity: 97.458  Percent Identity: 96.610

alignment_block:
US-09-327-750d-35 x BG070341/rev ..
Align seg 1/1 to reverse of: BG070341 from: 1 to: 747
1 MetAlaSerLys...LysGlnValIleLeuAspLeuThrValGluLysAs 16
|||||
581 ATGGCATCAAAATTTAAACAAGTATCATCTGGATCTCACTGGAGAAGA 532
|||||
16 pLysLysAspLysArgGlyGlyLysAlaSerLysGlnSerGluGluLup 33
|||||

```

531 CAAAAGACACAAAAAGCTGGAGGCTCCAAAACAAAGTGAAGAAGAAC 482

33 roHisHisLeuGluValGluAsnLysLysProGlyGlyAsnValArg 49

481 CCCACCATCTGGAAGAGGTTTGAACAAGAAGCTGGGGCAATGTCCGA 432

50 ArgLysValArgArgLeuValProAsnPhelLeuTrpAlaIleProAsnAr 66

431 AGGAAGTCAGCGCACTTGTGCTTAACCTTCTCTGGCCCATACCAATAG 382

66 gHisValAspArgAsnGluGlyGlyValGluAspValGlyArgPheValValG 83

381 GCATGTTGATGCATGAAGGGGAGAGGATGTTGGGAGATTGTAGTGC 332

83 InGlyThrGluValLysArgLysThrThrGluGlnGlnValArgProTyr 99

331 AGGGAACAGACAGTCAAGAGAAGACTACGGAGCAGCAGGTGACGCT 283

100 ArgArgPheArgThrProGluProAsnHisTyrAspPheCysLeuI 116

282 AGCGTTTCCGAACCCCGAACCTGACATCATTTACGACTTTGGCTCAT 233

116 ePro 117

232 ACCT 229

seq_name: gb_est2:BF723075

seq_documentation_block:

LOCUS BF723075 519 bp mRNA EST 03-JAN-2001

DEFINITION mab27e12.y1 Soares_NMEBA_branchial_arch Mus musculus cDNA clone

IMAGE:3971447 5', similar to TR:Q9Y516 Q9Y516 D3635G19.2.3 ;, mRNA

sequence.

ACCESSION BF723075

VERSION BF723075.1 GI:12024077

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 519)

AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL Tumor Gene Index

COMMENT Unpublished (1997)

Other_ESTs: mab27e12.x1

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo

, Ph.D.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CCAP clone distribution information can be

found through the I.M.A.G.E. Consortium/ILNL at:

image.ilnl.gov/image/html/iresources.shtml

MGI:1471479

Seq primer: -40RP from Gibco

High quality sequence stop: 472.

Location/Qualifiers

1..519

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:3971447"

/clone.lib="Soares_NMEBA_branchial_arch"

/tissue_type="branchial arches"

/dev_stage="embryo 10.5 dpc"

/lab_host="DH10B (phage resistant)"

/note="vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTGGGAGCGCGCATGCTTTTTTTTTTTTTTTT

3']; double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed and normalized by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 158 a 109 c 170 g 81 t 1 others
ORIGIN

alignment_scores: Quality: 568.50 Length: 113
Ratio: 5.168 Caps: 1
Percent Similarity: 97.345 Percent Identity: 96.460

alignment_block:
US-09-327-750D-35 x BF723075

Align seg 1/1 to: BF723075 from: 1 to: 519

1 MetAlaSerLys...LysGlnValIleLeuAspLeuThrValGluLysAs 16
181 ATGGCATCCAAATTTAAACAAGTCACTACCTGATCTGCTGGGAGAAAGA 230
16 pLysLysAspLysArgGlyGlyLysAlaSerLysGlnSerGluGluGluP 33
231 CAAAAAGACAAAAAGGTGGGAGGCTCCAAACAAAGTGAAGAAGAAC 280
33 roHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 49
281 CCCACCATCTGGAAGAGGTGAAACAAGAGCGCTGGGGAAATGTCCGA 330
50 ArgLysValArgArgLeuValProAsnPheLeuTIPAlaIleProAsnAr 66
331 AGGAAGTCAAGGACCTTGCCCTACTTCTCTGGCCATACCTAATAG 380
66 gHisValAspArgAsnGluGlyGlyGluAspValGlyArgPheValValG 83
381 GCATGTTGATCGCAATGAAGGGGAGAGGATGTTGGGAGATTGTAGTGC 430
83 InGlyThrGluValLysArgLysThrGluGlnGlnValArgProTyr 99
431 AGGAACAGANGTCAAGAGAAAGACTACCGAGCAGCGGTGAGGCTTAC 480
100 ArgArgPheArgThrProGluProAspAsnHisTyrAsp 112
481 AGGCGTTTCCAGCGCGGACCTGACAATCATTTACGAC 519
seq_name: gb_est1:BE333774

seq_documentation_block: 463 bp mRNA EST 14-JUL-2000
LOCUS BE333774
DEFINITION us27h06.y1 Soares_NMEBA_branchial_arch Mus musculus cDNA clone IMAGE:3168347 5' similar to SW:HG74_HUMAN O00994 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74. ; mRNA sequence.

ACCESSION BE333774
VERSION BE333774.1 GI:9207550
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 463)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
CDNA Library Email: nlnh.gov

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:

image.lnl.gov/image/html/iresources.shtml

MGI:1063807
Seq primer: -40RP from Gibco.

FEATURES
source

1. 463
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3168347"
/tissue.type="branchial_arch"
/dev_stage="embryo, 10.5 dpc"
/lab_host="DH10B (phage resistant)"
/note="Vector: pT73D-pac (Pharmacia) with a modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCATGCAATTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed and normalized by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 155 a 91 c 147 g 70 t
ORIGIN

alignment_scores: Quality: 544.50 Length: 113
Ratio: 4.950 Caps: 1
Percent Similarity: 97.345 Percent Identity: 95.575

alignment_block:
US-09-327-750D-35 x BE333774

Align seg 1/1 to: BE333774 from: 1 to: 463

1 MetAlaSerLys...LysGlnValIleLeuAspLeuThrValGluLysAs 16
126 ATGGCATCCAAATTTAAACAAGTCACTACCTGATCTGCTGGGAGAAAGA 175
16 pLysLysAspLysArgGlyGlyLysAlaSerLysGlnSerGluGluGluP 33
176 CAAAAAGACAAAAAGGTGGGAGGCTCCAAACAAAGTGAAGAAGAAC 225
33 roHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 49
226 CCCACCATCTGGAAGAGGTGAAACAAGAGCGCTGGGGAAATGTCCGA 275
50 ArgLysValArgArgLeuValProAsnPheLeuTIPAlaIleProAsnAr 66
276 AGGAAGTCAAGGACCTTGCCCTACTTCTCTGGCCATACCTAATAG 325
66 gHisValAspArgAsnGluGlyGlyGluAspValGlyArgPheValValG 83
326 GCATGTTGATCGCAATGAAGGGGAGAGGATGTTGGGAGATTGTGA GTCC 374
83 InGlyThrGluValLysArgLysThrGluGlnGlnValArgProTyr 99
375 AGGACAGACAGTCAAGAGAAAGACTACCGAGCAGCGGTGAGGCTTAC 424
100 ArgArgPheArgThrProGluProAspAsnHisTyrAsp 112
425 AAGCGTTTCCAGCGCGGACCTGACAATCATTTACGAC 463
seq_name: gb_est2:W17712

seq_documentation_block: 317 bp mRNA EST 10-SEP-1996
LOCUS W17712
DEFINITION md77604.r1 Soares mouse p3NM19.5 Mus musculus cDNA clone IMAGE:335454 5' similar to SW:HG74_HUMAN Q00994 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74. [1]; mRNA sequence.

ACCESSION W17712
VERSION W17712.1 GI:1292113

KEYWORDS SOURCE

EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 317)
Marra M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theisinger, B., Wyllie, T., Lennon, G., Soares, B., Wilton, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
WGI:216854

TITLE JOURNAL COMMENT

seq_name: gb_est1:AI227867
seq_documentation_block:
LOCUS AI227867 568 bp mRNA 20-JAN-1999
DEFINITION EST224562 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
RBC482 3' end, mRNA sequence.
ACCESSION AI227867
VERSION AI227867.1 GI:3811754
KEYWORDS EST
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

FEATURES source

Seq primer: mob.REGA+ET
High quality sequence stop: 291.
Location/Qualifiers
1..317
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:335454"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCATCTGAGTGGAGCGCGCATTTTCTTTTCTTTT 3'].
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."
BASE COUNT 108 a 63 c 87 g 59 t
ORIGIN

alignment_scores: Quality: 541.00 Length: 101 Ratio: 5.410 Gaps: 0 Percent Similarity: 99.010 Percent Identity: 98.020

alignment_block:
US-09-327-750D-35 x W17712 ..
Align seg 1/1 to: W17712 from: 1 to: 317

17 LysLysAspLysArgGlyGlyLysAlaSerLysGlnSerGluGluGluPr 33

1 AAAAAACAAAAAGGTGGAGGCTCCAAACAAAGTGAAGAGAAC 50

33 OHSHISLeuGluGluValGluAsnLysLysProGlyGlyAsnValArgA 50

51 CCACCATCTGGAAGAGGTGGAACAAAGAGCGCTGGGGAAATGTCCGAA 100

50 iGlyValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnArg 66

101 CCAAAAGTCAGCGACTTGTGCCTAATCTTCTGGGCCATACCAATAGG 150

67 HIsValAspArgAsnGluGlyGlyGluAspValIglYArgPheValIgl 83

151 CATGTGTGTCATGAAGGGGAGAGGATGTTGGGAGATTGTAGTGCA 200

83 nGlyThrGluValLysArgLysThrThrGluGlnGlnValArgProTyrA 100
|||||
201 GGAACAGAGTCAAGAGAAAGACTACGAGCAGCAGGTGAGGCTTACA 250
|||||
100 rGArgPheArgThrProGluProAsnHisTyrAspPheCysLeuIle 116
|||||
251 GCGGTTTCGAACCCCGGACACTGACATCATTCAGACTTTTGCCCTATA 300
|||||
117 Pro 117
|||
301 CCT 303

seq_name: gb_est1:AI227867

seq_documentation_block:

LOCUS AI227867 568 bp mRNA 20-JAN-1999
DEFINITION EST224562 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
RBC482 3' end, mRNA sequence.
ACCESSION AI227867
VERSION AI227867.1 GI:3811754
KEYWORDS EST
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE 1 (bases 1 to 568)

AUTHORS Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
Kerlavage, A.R. and Adams, M.D.
Rat genome Project: Generation of a Rat EST (REST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT Other ESTs: TC58086
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
1..568
/organism="Rattus sp."
/db_xref="ATCC (inhost):2036891"
/db_xref="taxon:10118"
/clone="RBC482"
/note="Organ: brain; Vector: pT73pac; Site_1: EcoRI;
Site_2: NotI"

BASE COUNT 140 a 143 c 108 g 177 t

ORIGIN

alignment_scores: Quality: 539.50 Length: 118 Ratio: 4.860 Gaps: 1 Percent Similarity: 94.068 Percent Identity: 87.288

alignment_block:
US-09-327-750D-35 x AI227867/rev ..
Align seg 1/1 to reverse of: AI227867 from: 1 to: 568

1 MetalaserLys...LysGlnValIleLeuAspLeuThrValGluLysAs 16

528 ATGGCGTCCAAAGTCAACAAAGTCACTACTGATCTCACTGCGAGAAAGA 479

16 PhysLysAspLysArgGlyGlyLysAlaSerLysGlnSerGluGluGluP 33

478 CAAGAAAAACAAAAAGGTGGAGGCTCCAAACAAAGTGAAGAGAAAT 429

33 roHISHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 49

```

|||||
428 CCCACCTCTGGAGAGGTTGAAACACAGAGGCTGGGGCAATGTCAGG 379
|||||
50 ArgLysValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnAr 66
|||||
378 AGGAAGTCAGGCGCACTGTGCTTAACCTTTCTATGGGCCATACCTAATAG 329
|||||
66 gHisValAspArgAsnGluGlyGlyGluAspValGlyArgPheValValG 83
|||||
328 GCATGTTGATCAGAGTGAAGGGGAGAGAGGTGGGAGATTCGTAGGCC 279
|||||
83 lnglyThrGluVallyArgLysThrThrGluGlnGlnValArgProTyr 99
|||||
278 AGGTGATGGAGGCAAGAGAAAGTCTAAGGAGCAACAGATGAGGCCCTTAC 229
|||||
100 ArgArgPheArgThrProGluLupProAsnHisTyrAspPheCysLeuI 116
|||||
228 ACGGTTTCCCAACCCCTGAACCTGACATCATTAGGACTTTTGCCTCAT 179
|||||
116 ePro 117
|||||
178 ACCT 175

```

seq_name: gb_est2:BI289546

```

seq_documentation_block: 643 bp mRNA EST 19-JUL-2001
LOCUS BI289546
DEFINITION UI-R-DKO-c12-0-UI.s1 UI-R-DKO Rattus norvegicus cDNA clone
UT-R-DKO-c12-0-UI 3', mRNA sequence.
ACCESSION BI289546
VERSION BI289546
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 643)
Bonalido,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized rat kidney pool library cDNA Library Preparation: M.B.
Soares Lab Clones distribution: clones will be available through
Research Genetics (www.resgen.com) The following repetitive
elements were found in this cDNA sequence: 1-51,
>Ar_rich#Low_complexity
Seq primer: M13 Forward
POLYA=Yes.
Location/Qualifiers
1..643
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/gb_xref="taxon:10116"
/clone="UI-R-DKO-c12-0-UI"
/clone_lib="UI-R-DKO"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT730-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DKO

```

FEATURES
source

library is a subtracted library derived from a mixture of five individually tagged normalized rat libraries: brain-nRBP (20%), heart-nRHP (20%), kidney-nRKP (20%), aorta-nRAP (20%), and placenta-nRPP (20%). Each original library was constructed from a mixture of equal amounts of RNA from seven different developmental time-points: embryonic day 17, embryonic day 19, embryonic day 21, embryonic day 1, adult day 12, adult day 75, and adult day 200. (Exception: the aorta pool does not contain embryonic day 17 RNA and the placenta pool contains only the three embryonic stages). Each library was normalized individually according to the procedure described by Bonaldo, Lennon & Soares (Genome Research Genome 6: 791-806, 1996). For construction of the DKO subtracted library, plasmid DNA from each of the five individually tagged normalized libraries was mixed in the proportions specified above and electroporated into competent bacteria for production of single-stranded circular DNA representing the pool of libraries. Single-stranded circular DNA representing these five normalized libraries was then used as a tracer in a subtractive hybridization with a driver (PCR amplified inserts from a plasmid DNA template preparation) comprising: a) a set of about 1,000 arrayed clones from each of the five non-normalized libraries of brain (CtOs), heart (CSOs), kidney (CUOs), aorta (CWOs), and placenta (CXOs). The resulting pool of approximately 5,000 clones represented about 33.3% of the final driver population. A set of about 2,000 arrayed clones from each of the five normalized libraries of brain (CTO), heart (CSO), kidney (CUO), aorta (CWO), and placenta (CXO). The resulting pool of about 10,000 clones represented about 66.6% of the final driver population.

BASE COUNT 160 a 149 c 113 g 221 t
ORIGIN

```

alignment_scores:
Quality: 511.50 Length: 118
Ratio: 4.608 Gaps: 1
Percent Similarity: 94.068 Percent Identity: 86.441

alignment_block:
US-09-327-750D-35 x BI289546/rev ..
Align seg 1/1 to reverse of: BI289546 from: 1 to: 643

1 MetAlaSerLys...LysGlnValIleLeuAspLeuThrValGluLysAs 16
|||||
601 ATGGCGTCCAAAGTCAACCAAGTCATATCTGATCTCACTGTGGAGAAGA 552
|||||
16 pLysLysAspLysArgGlyGlyLysAlaSerLysGlnSerGluGluGp 33
|||||
551 CAAGAAACAAACAAAAGGTGGGAG.GCCTCCAAACAAAGTGAAGAAGAT 503
|||||
33 roHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 49
|||||
502 CCCACCTCTGGAAGAGGTGTAACAAACAGAGCCTGGGGCAATGTCAGG 453
|||||
50 ArgLysValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnAr 66
|||||
452 AGGAAGTCAGGCGACTTGTGCTTAACCTTTCTATGGGCCATACCTAATAG 403
|||||
66 gHisValAspArgAsnGluGlyGlyGluAspValGlyArgPheValValG 83
|||||
402 GCATGTTGATCAGTGAAGGGGAGAGAGGTGGGAGATTCGTAGGCC 353
|||||
83 lnglyThrGluVallyArgLysThrThrGluGlnGlnValArgProTyr 99
|||||
352 AGGTGATGGAACCAAGAGAAAGTCTAAGGAGCAACAGATGAGGCCCTTAC 303

```

100 ArgArgPheArgThrProGluProAspAsnHisTyrAspPheCysLeu1116
 302 ACAGCTTTCGGAACCCCTGAACCTGACAAATCATTACGACTTTTGCCTCAT 253

116 ePro 117

252 ACCT 249

seq_name: gb_est1:A1117429

seq_documentation_block:

LOCUS A1117429 312 bp mRNA EST 15-NOV-1996
 DEFINITION mouse embryonic region Mus musculus cDNA
 clone IMAGE:538606 5', similar to SW:HG74_HUMAN Q00994 OVARIAN
 GRANULOSA CELL 13.0 KD PROTEIN HGR74. ; mRNA sequence.

ACCESSION A1117429

VERSION A1117429.1 GI:1672442

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT

Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:325342

Seq primer: -40ml3 ET

High quality sequence stop: 287.

Location/Qualifiers

FEATURES

source

1. 312

/organism="Mus musculus"

/strain="C57BL/6 x DBA"

/db_xref="taxon:10090"

/clone_lib="Beddington mouse embryonic region"

/sex="pooled"

/tissue_type="embryo"

/dev_stage="7.5dpc"

/lab_host="DH12S"

/note="Organ: whole embryo; Vector: pCMV-SPORT; Site: 1;
 SalI; Site: 2; NotI; Cloned unidirectionally. Primer:
 Oligo dr. Gastrulating embryos were collected at 7.5dpc
 from C57BL/6 x DBA matings, excluding embryos that had
 developed head folds and all extraembryonic tissues.

Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).

Referenced in Development 121, 2479-2489 (1995)"

BASE COUNT

ORIGIN

alignment_scores:

Quality: 503.50

Ratio: 5.086

Percent Similarity: 99.000

Percent Identity: 98.000

alignment_block:

US-09-327-750D-35 x A1117429

Align seg 1/1 to: A1117429 from: 1 to: 312

1 MetAlaSerLys...LysGlnValIleLeuAspLeuThrValGluLysAs 16
 112 ATGGCATCCAAATTTAAACAAGTCATCTGATCTGCTGGAAGAAGA 61
 16 pLysLysAspLysArgGlyGlyLysAlaSerLysGlnSerGluGlu 33
 62 CAAAAAGAGACAAAAGGGTGGGAGGCTCCAAACAAAGTGAAGAAG 111
 33 roHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnVal 49
 112 CCCACCATCTGGAAGAGGTTGAAACAAGAGAGCCCTGGGGAAATG 161
 50 ArgLysValArgLysGluValProAsnPheLeuThrPalalilePro 66
 162 AGGAAAGTCAGCGACGCTTGTGCCTAACTTCTCTGGCCATACCTA 211
 66 gHisValAspArgAsnGluGlyGlyLysValGlyValGlyArgPhe 83
 212 GCATGTTGATCCCATGAAGGGGAGAGATGTTGGGAGATTTGTAG 261
 83 lngLysThrGluValLysArgLysThrGluGlnGlnValArgProTyr 99
 262 AGGACACAGAGTCAAGAGAGAGAGTACGAGCAGCAGAGGTGAGC 311
 seq_name: gb_est2:W35893

seq_documentation_block:

LOCUS W35893 472 bp mRNA EST 14-MAY-1996
 DEFINITION mc53g09.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
 clone IMAGE:352288 5', similar to SW:HG74_HUMAN Q00994 OVARIAN
 GRANULOSA CELL 13.0 KD PROTEIN HGR74. [1] ; mRNA sequence.

ACCESSION W35893

VERSION W35893.1 GI:1317733

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 472)

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

Waterston, R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:224088

Seq primer: ETPrimer

High quality sequence stop: 441.

Location/Qualifiers

1. 472

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_lib="IMAGE:352288"

/sex="unknown"

/tissue_type="embryo"

/dev_stage="13.5-14.5dpc total fetus"

/lab_host="DH10B"

/note="Vector: pT73D-pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5,

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 465)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Willson, R. and
Waterston, R.

The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:229324

Seq primer: EYPrimer

High quality sequence stop: 328.

Location/Qualifiers

1. .465

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_image="367892"

/clone_lib="Soares mouse embryo NbME13.5 14.5"

/sex="unknown"

/tissue_type="embryo"

/dev_stage="13.5-14.5dpc total fetus"

/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TCGTACCAATCTGAGTGGAGCGCGCGGAATTTTCTTTTCTTTT
1 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."

BASE COUNT 147 a 90 c 156 g 72 t

alignment_scores:

Quality: 469.50 Length: 99
Ratio: 4.840 Gaps: 1
Percent Similarity: 97.980 Percent Identity: 96.970

alignment_block:

US-09-327-750D-35 x W54487

Align seg 1/1 to: W54487 from: 1 to: 465

1 MetAlaSerLys...LysGlnValIleLeuAspLeuThrValGluLysAs 16
|||||
168 ATGGCATCCAAATTTAAACAAGTCATCTGATCTCCTGCTGGAGAAAGA 217
|||||
16 pLysLysAspLysArgGlyGlyLysAlaSerLysGlnSerGluGluGlup 33
|||||
218 CAAAAAGACAAAAAGAGTGGGAAGGCCCTCCAAACAAGTGAAGAAGAAC 267
|||||
33 roHisLysLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 49
|||||
268 CCCACCATCTGGAGAGGTTGAACAACAGACCCCTGGGGGAATGTCCGA 317
|||||
50 ArgLysValArgArgLeuValProAsnPhleLeuTrpAlaIleProAsnAr 66
|||||

318 AGAAAGTC..AGCGACTTGTGCTAACTTTCTCTGGGCCATACCAATAG 366
66 gHisValAspArgAsnGluGlyGlyGluAspValGlyArgPheValValG 83
|||||
367 GCATGTTGATCCATGAAGGGGAGAGGATGTTGGGAGATTTGTAGTGC 416
83 InGlyThrGluValLysArgLysThrThrGluGlnGlnValArgPro 98
|||||
417 AGGGAACAAGCAAGTCAAGAGAAGACTAGCGAGCAGCAGGTGAGGCCT 463

seq_name: gb_est1:AL118340

seq_documentation_block:

LOCUS AL118340 422 bp mRNA EST 23-SEP-1999
DEFINITION v911326 Beddington mouse dissected endoderm Mus musculus CDNA
clone 528_13N20 5', mRNA sequence.

ACCESSION AL118340

VERSION AL118340.1 GI:5920179

KEYWORDS EST_1

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 422)
Harrison, S.M., Dunwoodie, S.L., Arkell, R.M., Lehrach, H. and
Beddington, R.S.

Isolation of novel tissue-specific genes from cDNA libraries
representing the individual tissue constituents of the gastrulating
mouse embryo

Development 121 (8), 2479-2489 (1995)

95401865

Contact: Ruiz, P., Lehrach, H. and Avner, P.

ESC Mouse Transcription Mapping Consortium

Genoscope - CNS

2, rue Gaston Cremieux, CP 5706, 91057 Evry CEDEX, France

Email: pavner@pasteur.fr

Clone available from Ressourcenzentrum, Heubnerweg 6, D-14059

Berlin, Germany. Web site http://www.rzpd.de

Seq primer: CGGTCGGGAATTCCTCCGGT.

Location/Qualifiers

1. .422

/organism="Mus musculus"

/strain="C57BL6 x DBA"

/db_xref="taxon:10090"

/clone="528_13N20"

/tissue_type="dissected endoderm"

/dev_stage="7.5 dpc"

/note="Vector: pSport1; Site 1: NotI; Site 2: SalI; Cloned
unidirectionally. - High quality sequence only submitted. -
Average insert size: 1.2 kb (range: 0.2 - 2.kb)"

BASE COUNT 128 a 86 c 143 g 65 t

ORIGIN

alignment_scores:

Quality: 463.50 Length: 92
Ratio: 5.093 Gaps: 1
Percent Similarity: 98.913 Percent Identity: 98.913

alignment_block:

US-09-327-750D-35 x AL118340

Align seg 1/1 to: AL118340 from: 1 to: 422

1 MetAlaSerLys...LysGlnValIleLeuAspLeuThrValGluLysAs 16
|||||
146 ATGGCATCCAAATTTAAACAAGTCATCTGATCTCCTGCTGGAGAAAGA 195
|||||
16 pLysLysAspLysArgGlyGlyLysAlaSerLysGlnSerGluGluGlup 33
|||||
196 CAAAAAGACAAAAAGAGTGGGAAGGCCCTCCAAACAAGTGAAGAAGAAC 245
|||||

33 ROHSHISLeuGluValGluAsnLysLysProGlyGlyAsnValArg 49
|||||
246 CCCACCATCTGGAGAGGTTGAACAAGAGCCCTGGGGGAATGTCCGA 295
|||||
50 ArgLysValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnAr 66
|||||
296 AGGAAAGTCAGCGGACTTCTGCTTAACCTCTCTGGGCCATACCAATAG 345
|||||
66 gHisValAspArgAsnGluGlyGlyValGluAspValGlyArgPheValValG 83
|||||
346 GCATGTGTGATCGCAATGAGGGGGGAGAGATGTTGGGAGATTCTCGTGC 395
|||||
83 InGlyThrGluValLysArgLysThr 91
|||||
396 AGGGAACAGAGTCAAGAGAAAGACT 421
|||||

seq_name: gb_est1.AW251884

seq_documentation_block: 540 bp mRNA EST 17-DEC-1999
LOCUS AW251884
DEFINITION UI-R-BJO-adn-b-06-0-UI.s1 UI-R-BJO Rattus norvegicus cDNA clone
UI-R-BJO-adn-b-06-0-UI 3', mRNA sequence.

ACCESSION AW251884
VERSION AW251884.1 GI:6595475
KEYWORDS EST.

SOURCE Norway rat.
Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 540)
AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene
discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477

COMMENT Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mscores@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized atrium at 16.5 dpc library cDNA Library Preparation:
M.B. Soares Lab Clone distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source

Location/Qualifiers
1..540

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJO-adn-b-06-0-UI"
/clone_lib="UI-R-BJO"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pMT3D-Pac (Pharmacia) with a modified
polylinker; Site1: Not I; Site2: Eco RI; The UI-R-BJO
library is a subtracted library derived from the UI-R-AAL,
UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and
UI-R-AG1 libraries. These libraries represent tissues from
rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal
at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
canal at 15 dpc, and ventricle at 13 dpc. The tag is a
string of 5-6 nucleotides present between the Not I site
and the oligo-dT track. The library was constructed as
described by Ronaldo, Lennon and Soares, Genome Research

6: 791-806, 1996.
TAG_LIB=UI-R-BJO
TAG_TISSUE=atrium at 16.5 dpc
TAG_SEQ=GATTC.
BASE COUNT 146 a 120 c 94 g 180 t
ORIGIN
alignment_scores:
Quality: 459.00 Length: 97
Ratio: 5.044 Gaps: 0
Percent Similarity: 93.814 Percent Identity: 86.598
alignment_block:
US-09-327-750D-35 x AW251884/rev ..
Align seg 1/1 to reverse of: AW251884 from: 1 to: 540

21 ArgGlyGlyLysAlaSerLysGlnSerGluGluProHisHisLeuG 37
:::|||||
538 AAAGGTGGGAGGCGCTCCAAACAAAGTGAAGAGGAATCCCACTCTGGA 489
|||||
37 uGluValGluAsnLysLysProGlyGlyAsnValArgArgLysValArgA 54
|||||
488 AGAGGTTGAAACAAAGAAAGCCTGGGGCAATGTCAGGAGGAAAGTCAGGC 439
|||||
54 rgLeuValProAsnPheLeuTrpAlaIleProAsnArgHisValAspArg 70
|||||
438 GACTGTGCTTAACCTTTCTATGGGCCATACCTAATAGGCATGTTGATCAC 389
|||||
71 AsnGluGlyGlyGluAspValGlyArgPheValValGlnGlyThrGluVa 87
:::|||||
388 AGTGAAGGGGAGAGGAGGTTGGGAGATTCGTAGGCGCAGGTGATGGAAGC 339
|||||
87 lLysArgLysThrThrGluGlnGlnValArgProTyrArgArgPheArgT 104
:|||||
338 CAAGAGAAAGTCTAAGGAGCAACAGATGAGGCCTTACAGCGCTTTCCGAA 289
|||||
104 hrProGluProAsnAsnHisTyrAspPheCysLeuIlePro 117
|||||
288 CCCCTGAACCTGACAAATCATTACGACTTTTGCCTCATCACT 248
|||||

OM of: US-09-327-750D-35 to: N_Geneseq_1101.* out_format : pfs
 Date: Mar 11, 2002 3:42 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODELframet_p2n.model -DEV=xlh
 -Q/c9n2_1/USPTO.spool/US09327750/runat_11032002_101154_20362/app_query.fasta_1.1472
 -DB=N_Geneseq_1101 -QFMT=fastp -SUFFIX=ring -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
 -TRANS=human40 cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09327750_@CGN1_1_330 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-327-750D-35
 Query length: 117
 Database: N_Geneseq_1101.*
 Database sequences: 930621
 Database length: 428662619
 Search time (sec): 355.560000

score_list:

Sequence	Strd Orig	Zscore	Escore	Len	Documentation
/SID2/gcgdata/geneseq/NA2001.DAT:AAH03517	+	355.50	767.81	1.1e-34	865
/SID2/gcgdata/geneseq/NA2001.DAT:AAH13750	+	355.50	764.27	1.7e-34	1229
/SID2/gcgdata/geneseq/NA2001.DAT:AAH03548	+	355.50	763.22	2.0e-34	1364
/SID2/gcgdata/geneseq/NA2001.DAT:AAH23529	+	175.00	367.98	2.0e-12	891
/SID2/gcgdata/geneseq/NA2000.DAT:AAH21748	+	174.00	365.47	2.8e-12	917
/SID2/gcgdata/geneseq/NA2001.DAT:AAH23328	+	153.00	324.71	7.0e-10	700
/SID2/gcgdata/geneseq/NA2000.DAT:AAH01005	+	147.00	312.12	2.6e-09	485
/SID2/gcgdata/geneseq/NA2000.DAT:AAH01005	+	141.00	297.91	1.6e-08	532
/SID2/gcgdata/geneseq/NA2001.DAT:AAH75810	+	99.50	202.37	0.0034	767
/SID2/gcgdata/geneseq/NA2001.DAT:AAH15881	+	99.50	201.19	0.0039	862
/SID2/gcgdata/geneseq/NA2001.DAT:AAH59611	+	99.50	200.78	0.0042	898
/SID2/gcgdata/geneseq/NA1999.DAT:AAH23519	+	98.50	159.28	0.8526	4453
/SID2/gcgdata/geneseq/NA2001.DAT:AAH45143	+	94.00	189.87	0.0169	792
/SID2/gcgdata/geneseq/NA2000.DAT:AAH01003	+	92.00	182.29	0.0124	401
/SID2/gcgdata/geneseq/NA2001.DAT:AAH20201	+	89.00	189.41	0.0179	276
/SID2/gcgdata/geneseq/NA2001.DAT:AAH45401	+	89.00	189.41	0.0179	276
/SID2/gcgdata/geneseq/NA2001.DAT:AAH05905	+	89.00	189.41	0.0179	276
/SID2/gcgdata/geneseq/NA2001.DAT:AAH13308	+	87.50	181.99	0.0463	415
/SID2/gcgdata/geneseq/NA2001.DAT:AAH13461	+	87.50	181.99	0.0463	415
/SID2/gcgdata/geneseq/NA2001.DAT:AAH03209	+	87.50	181.99	0.0463	415
/SID2/gcgdata/geneseq/NA2000.DAT:AAH03880	+	87.50	177.29	0.0847	662
/SID2/gcgdata/geneseq/NA2000.DAT:AAH03879	+	87.50	176.75	0.0907	698
/SID2/gcgdata/geneseq/NA2001.DAT:AAH160367	+	87.50	174.68	0.1184	858
/SID2/gcgdata/geneseq/NA2000.DAT:AAH10889	+	84.50	170.20	0.2101	692
/SID2/gcgdata/geneseq/NA2000.DAT:AAH70191	+	82.00	157.98	1.01	1344
/SID2/gcgdata/geneseq/NA1990.DAT:AAH06000	+	81.50	150.50	2.63	2533
/SID2/gcgdata/geneseq/NA2000.DAT:AAH64139	+	80.50	118.25	164.41	50000
/SID2/gcgdata/geneseq/NA1998.DAT:AAH21511	+	80.00	137.30	14.29	6755
/SID2/gcgdata/geneseq/NA2001.DAT:AAH71145	+	79.50	149.29	3.07	1839
/SID2/gcgdata/geneseq/NA2000.DAT:AAH73711	+	77.50	141.89	7.93	2471
/SID2/gcgdata/geneseq/NA2001.DAT:AAH13331	+	77.50	134.33	75.44	14141
/SID2/gcgdata/geneseq/NA2001.DAT:AAH29749	+	76.00	133.90	22.10	3929
/SID2/gcgdata/geneseq/NA1999.DAT:AAH13264	+	76.00	129.63	38.21	6005
/SID2/gcgdata/geneseq/NA1999.DAT:AAH231091	+	75.50	144.44	5.72	1236
/SID2/gcgdata/geneseq/NA1990.DAT:AAH06613	+	75.50	129.01	41.37	5721
/SID2/gcgdata/geneseq/NA1989.DAT:AAH91167	+	75.50	127.91	47.63	6381
/SID2/gcgdata/geneseq/NA2001.DAT:AAH22893	+	75.00	143.96	6.08	1161
/SID2/gcgdata/geneseq/NA2000.DAT:AAH38803	+	74.50	141.40	8.45	1342
/SID2/gcgdata/geneseq/NA2000.DAT:AAH49751	+	74.50	141.39	8.46	1343
/SID2/gcgdata/geneseq/NA1997.DAT:AAH12788	+	74.50	127.88	47.82	5137

/SID2/gcgdata/geneseq/NA1999.DAT:AAZ01022 - 74.50 103.74 1.1e+03 5652
 /SID2/gcgdata/geneseq/NA2001.DAT:AAI21493 + 74.00 152.63 2.00 394
 /SID2/gcgdata/geneseq/NA2001.DAT:AAI46784 + 74.00 152.63 2.00 394
 /SID2/gcgdata/geneseq/NA2001.DAT:AAI07190 + 74.00 152.63 2.00 394
 /SID2/gcgdata/geneseq/NA1995.DAT:AAQ86237 + 74.00 128.11 46.45 4500

seq_name: /SID2/gcgdata/geneseq/NA2001.DAT:AAH03517

seq_documentation_block:

ID AAH03517 standard; cDNA: 865 BP.
 XX
 AC AAH03517;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (5'-primer) SEQ ID NO:352.
 XX
 KW Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PR 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 XX
 PR 27-AUG-1999; 99JP-0300253.
 XX
 PR 11-JAN-2000; 2000JP-0118776.
 XX
 PR 02-MAY-2000; 2000JP-0183767.
 XX
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 (HELI-) HELIX RES INST.
 PA
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 full-length cDNAs defined in the specification, and for the detection
 and/or diagnosis of the abnormality of the proteins encoded by the
 full-length cDNAs -
 XX
 PS Claim 1; SEQ ID 352; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 full-length cDNAs defined in the specification. Where a primer set
 comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 to the complementary strand of a polynucleotide which comprises one of
 the 5602 nucleotide sequences defined in the specification, where the
 oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 of an oligonucleotide comprising a sequence complementary to the
 complementary strand of a polynucleotide which comprises a 5'-end
 sequence and an oligonucleotide comprising a sequence complementary to a
 polynucleotide which comprises a 3'-end sequence, where the
 oligonucleotide comprises at least 15 nucleotides and the combination of
 the 5'-end sequence/3'-end sequence is selected from those defined in
 the specification. The primer sets can be used in antisense therapy and
 in gene therapy. The primers are useful for synthesizing polynucleotides,
 particularly full-length cDNAs. The primers are also useful for the
 detection and/or diagnosis of the abnormality of the proteins encoded by
 the full-length cDNAs. The primers allow obtaining of the full-length
 cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 represent oligonucleotides, all of which are used in the exemplification
 of the present invention.

Sequence 865 BP; 239 A; 177 C; 224 G; 221 T; 4 other;

alignment_scores:
 Quality: 355.50 Length: 120
 Ratio: 3.665 Gaps: 3
 Percent Similarity: 80.833 Percent Identity: 59.167

alignment_block:
 US-09-327-750D-35 x AAH03517

Align seg 1/1 to: AAH03517 from: 1 to: 865

```

1 MetAlaSerLysLysGlnValIleLeu...AspLeuThrValGluLysAs 16
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:
191 ATGGAGTCCAAAGAGGAACTAGCGCAACATCTCAACGGGGAATGC 240

16 pLysLysAspLysArgGlyGly...LysAlaSerLysGlnSerGluGlu 32
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:
241 CCAACAAGAAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 290

32 LuProHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnVal 48
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:
291 AATCCCGCCATTTGGGAGGGGGTGAAGGCCAGAACCTGGAGGAATATC 340

49 ArgArg...LysValArgArgLeuValProAsnPheLeuTirPAlaIlePr 64
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:
341 AGCGGGGGGCGAGTTAGCGGACTGTCTTAATTTTCGATGGCCATACC 390

64 oAsnArgHisValAspArgAsnGluGlyGluAspValGlyArgPheV 81
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:
391 TANTAGCATATTGAGCACAATGACGAGAGATGATGTAGAAAGTTTG 440

81 alValGlnGlyThrGluValLysArgLysThrThrGluGlnValArg 97
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:
441 TAGGCAGATGATGGAATCAAGAGAAAGACTAGGGAACAGCAGATGAGG 490

98 ProTyrArgPheArgThrProGluProAspAsnHisTyrAspPheCy 114
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:
491 CACTATATGCGCTTCCAACTCTTGAACTCTGACCTGACCACTTAATGACTTTTG 540

114 sLeuIlePro 117
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:
541 CCTCATACCT 550

```

seq_name: /SIDS2/gcdata/geneseq/geneseq/NA2001.DAT:AAH13750

seq_documentation_block:

XX AAH13750 standard; cDNA; 1229 BP.

AC AAH13750;

DT 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:10656.

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

OS Homo sapiens.

XX EPI074617-A2.

PD 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

PI

XX WPI: 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

XX full-length cDNAs.

PS Claim 8; SEQ ID 10656; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification, where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to

CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX Sequence 1229 BP; 351 A; 243 C; 296 G; 339 T; 0 other;

alignment_scores:
 Quality: 355.50 Length: 120
 Ratio: 3.665 Gaps: 3
 Percent Similarity: 80.833 Percent Identity: 59.167

alignment_block:

US-09-327-750D-35 x AAH13750

Align seg 1/1 to: AAH13750 from: 1 to: 1229

```

1 MetAlaSerLysLysGlnValIleLeu...AspLeuThrValGluLysAs 16
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:
191 ATGGAGTCCAAAGAGGAACTAGCGCAACATCTCAACGGGGAATGC 240

16 pLysLysAspLysArgGlyGly...LysAlaSerLysGlnSerGluGlu 32
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:
241 CCAACAAGAAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 290

32 LuProHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnVal 48
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:
291 AATCCCGCCATTTGGGAGGGGGTGAAGGCCAGAACCTGGAGGAATATC 340

49 ArgArg...LysValArgArgLeuValProAsnPheLeuTirPAlaIlePr 64
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:
341 AGCGGGGGGCGAGTTAGCGGACTGTCTTAATTTTCGATGGCCATACC 390

64 oAsnArgHisValAspArgAsnGluGlyGlyGluAspValGlyArgPheV 81
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:
391 TANTAGCATATTGAGCACAATGACGAGAGATGATGTAGAAAGTTTG 440

81 alValGlnGlyThrGluValLysArgLysThrThrGluGlnValArg 97
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:
441 TAGGCAGATGATGGAATCAAGAGAAAGACTAGGGAACAGCAGATGAGG 490

98 ProTyrArgPheArgThrProGluProAspAsnHisTyrAspPheCy 114
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:
491 CACTATATGCGCTTCCAACTCTTGAACTCTGACCTGACCACTTAATGACTTTTG 540

```


XX 17-OCT-2001 (first entry)
XX Human X chromosome linked gene expression protein 14 coding sequence.
DE Human; X chromosome linked gene expression protein 14; cancer;
XX HIV infection; cytostatic; anti-HIV; chromosome X; ss.
XX Homo sapiens.
XX CN1296969-A.
XX 30-MAY-2001.
XX 23-NOV-1999; 99CN-0124078.
XX 23-NOV-1999; 99CN-0124078.
XX (SHAN-) SHANGHAI BORONG GENE DEV CO LTD.
XX Mao Y, Xie Y;
PI WPI: 2001-483897/53.
DR P-PSDB; AAG66407.
XX Polypeptide-human X chromosome linked gene expression protein 14 and
PT polynucleotide for coding said polypeptide -
XX Claim 6; Page 24 (Disclosure); 3lpp; Chinese.
XX The present sequence is the coding sequence for human X chromosome linked
CC gene expression protein 14. The protein and coding sequence are useful
CC for treating diseases e.g. cancer and HIV infection.
XX Sequence 767 BP; 224 A; 143 C; 215 G; 185 T; 0 other;
SQ

alignment_scores:
Quality: 99.50 Length: 132
Ratio: 1.363 Gaps: 6
Percent Similarity: 55.303 Percent Identity: 29.545

alignment_block:
US-09-327-750D-35 x AAH75810 ..
Align seg 1/1 to: AAH75810 from: 1 to: 767

1 MetAlaSerLysLysGln.....ValIleLeuAspLe 11
101 ATGAGCTCCAAAGAGACGACGCTTAACAATCTCATCTGCGGAAATGT 150
11 uThrValGluLysAspLysLysAspLysArgGlyGlyLysAlaSerLysG 28
151 CAACACAGGAAATGATGAAAAAGATGAAAGAGGAGGAGCAAGTCTGCTAATAA. 199
28 lnSerGluGluGluProHisHisLeuGluGluValGluAspLysLysPro 44
200 ..GGGAGGCCCTGGCCCTACCTTGAATGTAGTATAGTACTGTGTCCT 247
45 GlyGlyAsnValArg...LysValArgArgLeuValProAsnPheLe 60
248 AGAGAAACCGTAGGCGGTTCGCGTTAGGAGGCCCATCTCGCAGTATAG 297
60 uTPtLalIleProAsnArgHisValAsp.....ArgAsnG 72
298 ATGGGACATAATGCATAGCTTGGAGCCACAGGACGAGGATGAGAGG 347
72 luGlyGlyGluAspValGlyArgPheValValGlnGlyThrGluValLys 88
348 AGAATATGAAAGGATGGGAGGAGGTGAGACAG.....CTGATG 388
89 ArgLysThrThrGluGlnValArgProTyrArgArgPheArgThrPr 105

389 GAAAAGCTGAGGAGGAAAGCAGTTGAGTCATAGTCTCGGGCAGTCAGCAC 438
105 oGluPro.....AspAsnHisTyrAspPheCysLeuIlePro 117
439 TGATCCCCCTCACCATGACCATCAGTATGAGTTTGCCTTATGCCC 484
seq_name: /SIDS7/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI58581
seq_documentation_block:
ID AAI58581 standard; cDNA; 862 BP.
XX AAI58581;
XX 22-OCT-2001 (first entry)
XX Human polynucleotide SEQ ID NO 784.
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX Homo sapiens.
XX WO200153312-A1.
XX 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US34263.
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0596042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
DR WPI: 2001-442253/47.
DR P-PSDB; AAM39425.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX Claim 1; SEQ ID NO 784; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX Sequence 862 BP; 235 A; 183 C; 238 G; 206 T; 0 other;
SQ

```

alignment_scores:
  Quality: 99.50      Length: 132
  Ratio: 1.363        Gaps: 6
  Percent Similarity: 55.303  Percent Identity: 29.545

alignment_block:
US-09-327-750D-35 x AAF58581 ..

Align seg 1/1 to: AAF58581 from: 1 to: 862
1 MetAlaSerLysLysGln.....ValileLeuAspLe 11
226 ATGGAGTCCAAAGAGCAAGCGCTTAAACAATCTCATCGTGGAATAATGT 275
11 uThrValGluLysAspLysLysAspLysArgGlyGlyLysAlaSerLysG 28
276 CAACCAAGNAATGATGAAAGAGTAAAGAGGACCAAGTGTCTAATAAA. 324
28 InSerGluGluGluProHisHisLeuGluGluValGluAsnLysLysPro 44
325 ..GGGAGGCCCTTGGCCCTACCTTTGAATGTTAGTGAATACTGTGTGCCT 372
45 GlyGlyAsnValArgArg...LysValArgArgLeuValProAsnPheLe 60
373 AGAGGNAACCGTAGCGGTCGCGGTAGGACGCCCACTCTGCAGTATAG 422
60 uTrpAlaIleProAsnArgHisValAsp.....ArgAsnG 72
423 ATGGGACATAATGATAGCTTGGAGAGCCACAGCAAGGATGAGAGAG 472
72 LuGlyGlyGluAspValGlyArgPheValGlnGlyThrGluValLys 88
473 AGATATGGAAGGATGGGAGAGGTGAGACAG.....CTGATG 513
89 ArgLysThrThrGluGlnValArgProTyrArgArgPheArgThrPr 105
514 GAAAGCTGAGGGAAGACAGTTGATCATAGTTTGGCGGCAGTCAGCAC 563
105 oGluPro.....AspAsnHisTyrAspPheCysLeuIlePro 117
564 TGATCCCTCCATCCATGACCATCAGGATGTTTGGCTTATGCC 609
seq_name: /SID52/9c9data/geneseq/geneseqn/NA2001.DAT:AAF59611
seq_documentation_block:
ID AAF59611 standard; cDNA: 898 BP.
AC AAF59611;
XX
XX
XX 24-APR-2001 (first entry)
XX
XX Human cell cycle and proliferation protein CCYPR-22 cDNA, SEQ ID NO:76.
XX
XX Cell cycle and proliferation protein; CCYPR; human; agonist;
KW antagonist; gene therapy; detection; gene therapy;
KW transgenic animal disease model; immune disorder;
KW developmental disorder; cell signalling disorder;
KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;
KW arteriosclerosis; asthma; allergy; diabetes mellitus;
KW menstrual cycle disorder; bacterial infection; ss.
XX
OS Homo sapiens.
XX
XX WO200107471-A2.
XX
XX 01-FEB-2001.
XX
XX 21-JUL-2000; 2000WO-US19948.
XX
XX 21-JUL-1999; 99US-0145075.
XX
XX 08-SEP-1999; 99US-0153129.
XX
PR
XX
XX 10-NOV-1999; 99US-0164647.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;
XX Azimzal Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;
XX
XX WPI; 2001-112727/12.
XX
XX P-PSDB; AAB60474.
XX
XX Human cell cycle and proliferation proteins and polynucleotides are
XX used to treat, diagnose and prevent immune, developmental and cell
XX signaling disorders and cell proliferative disorders including cancer.
XX
XX Claim 5; Page 181-182; 205pp; English.
XX
XX Sequences AAF59590-AAF59643 represent cDNAs encoding 54 human
XX cell cycle and proliferation proteins (CCYPR), AAB60453-AAB60506.
XX CCYPR and agonists of CCYPR are used to treat diseases or conditions
XX associated with decreased expression of functional CCYPR, while CCYPR
XX antagonists are used to treat diseases or conditions associated with
XX overexpression of functional CCYPR. Monoclonal or polyclonal antibodies
XX to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or
XX radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect
XX compounds e.g., antibodies, oligonucleotides and proteins (receptors)
XX that specifically bind to CCYPR, and in drug screening methods to
XX identify compounds that modulate the activity of CCYPR. CCYPR
XX nucleotides can be used to generate transgenic animal models of human
XX disease, and can be used in gene therapy in target cells with genetic
XX abnormalities with respect to the expression of CCYPR for the
XX treatment or prevention of a disorder associated with CCYPR.
XX Diseases which can be diagnosed, treated and prevented using CCYPR
XX proteins, nucleic acids, agonists or antagonists include immune,
XX developmental and cell signalling disorders, and cell proliferative
XX disorders including cancer. Specific examples of these disorders
XX include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
XX diabetes mellitus, disorders of the menstrual cycle and infections
XX caused by bacteria.
XX
XX Sequence 898 BP; 250 A; 186 C; 251 G; 211 T; 0 other;
XX
alignment_scores:
  Quality: 99.50      Length: 132
  Ratio: 1.363        Gaps: 6
  Percent Similarity: 55.303  Percent Identity: 29.545

alignment_block:
US-09-327-750D-35 x AAF59611 ..

Align seg 1/1 to: AAF59611 from: 1 to: 898
1 MetAlaSerLysLysGln.....ValileLeuAspLe 11
254 ATGGAGTCCAAAGAGCAAGCGCTTAAACAATCTCATCGTGGAATAATGT 303
11 uThrValGluLysAspLysLysAspLysArgGlyGlyLysAlaSerLysG 28
304 CAACCAAGNAATGATGAAAGAGTAAAGAGGACCAAGTGTCTAATAAA. 352
28 InSerGluGluGluProHisHisLeuGluGluValGluAsnLysLysPro 44
353 ..GGGAGGCCCTTGGCCCTACCTTTGAATGTTAGTGAATACTGTGTGCCT 400
45 GlyGlyAsnValArgArg...LysValArgArgLeuValProAsnPheLe 60
401 AGAGGNAACCGTAGCGGTCGCGGTAGGACGCCCACTCTGCAGTATAG 450
60 uTrpAlaIleProAsnArgHisValAsp.....ArgAsnG 72
451 ATGGGACATAATGATAGCTTGGAGAGCCACAGCAAGGATGAGAGAGG 500
72 LuGlyGlyGluAspValGlyArgPheValGlnGlyThrGluValLys 88

```

```
||||| ||| ..... ||| ||| ||| .....
501 AGAATATGGAAGATTGGGAGGAGTGACAG.....CTGATG 541
89 ArgLysThrThrGluGlnValArgProTyrArgPheArgThrPr 105
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
542 GAAAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 591
105 oGluPro.....AspAsnHisTyrAspPheCysLeuIlePro 117
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
592 TGATCCCTCCACCATGACCATGACGATGAGTTTGCCTTATGCC 637
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:AAx23519

seq_documentation_block:
ID AAX23519 standard; DNA; 44453 BP.

AC AAX23519;
DT 23-JUN-1999 (first entry)
DE Human kidney aminopeptidase P genomic DNA fragment 3.

XX Aminopeptidase: human; Amp; gene therapy; treatment; Amp-deficiency;
KW prenatal diagnosis; angiodema; antihypertensive agent; atherosclerosis;
KW arterial stenosis; industrial protein feed; malabsorption syndrome;
KW proteinaceous waste degradation; additive; immunohistochemistry; ss.

XX Homo sapiens.

OS Wo9911799-A2.

PN 11-MAR-1999.

XX 02-SEP-1998; 98WO-US18426.

XX 02-SEP-1997; 97US-0057854.

XX (MEDI-) MEDICAL COLLEGE GEORGIA RES INST.

XX Ryan JW, Sprinkle TJC, Venema RC;

XX WPI; 1999-205193/17.

XX Nucleic acid encoding human aminopeptidase P

PS Claim 13; Page 139-165; 201pp; English.

CC This invention describes the isolation of a novel human aminopeptidase P
CC (Amp). This protein is used to produce recombinant Amp and can be used
CC for gene therapy for treating Amp-deficiency conditions. Its fragments
CC are used as primers and probes to identify patients with homozygous and
CC heterozygous Amp deficiency, including prenatal diagnosis (patients
CC defective in Amp are at risk of developing angiodema if treated with
CC angiotensin-converting enzyme inhibitors), also as antisense inhibitors
CC in cases of excessive Amp expression. The product of the invention is
CC also used to identify Amp-expressing sequences in other animals and to
CC generate transgenic animals, and comparisons of genomic sequences are
CC used to detect mutations. Amp inhibitors are potentially useful as
CC antihypertensive agents and to prevent or treat arterial (re)stenosis
CC or atherosclerosis. The structure of Amp is used to design synthetic
CC substrates, e.g. for use in Amp assays. Amp, which hydrolyzes N-terminal
CC imido bonds, can be used to degrade industrial protein feeds to free
CC amino acids, to degrade proteinaceous wastes, as additives in enzyme
CC formulations used to treat malabsorption syndrome and for studying its
CC biological role. Antibodies against Amp are used in immunohistochemical
CC methods to study Amp distribution.

SQ Sequence 44453 BP; 13034 A; 10110 C; 10292 G; 11017 T; 0 other;

alignment_scores:
Quality: 98.50 Length: 102
Ratio: 1.539 Gaps: 2

Percent Similarity: 62.745 Percent Identity: 26.471

alignment_block:

US-09-327-750D-35 x AAX23519

Align seg 1/1 to: AAX23519 from: 1 to: 44453

```
1 MetAlaSerLysLysGlnValIleLeuAspLeuThrValGluLysAspLys 17
   ..... ||||| .....
43713 CTACAAGAAGAGAGGACACATGGCTTTTCAGTCAGCTTGGAGGAGAGAG 43762
   ..... ||||| .....
43763 GGAGGAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 43812
   ..... ||||| .....
34 1sHisLeuGluGluValGluAsnLysLysProGly..GlyAsnValArgA 50
   ||||| ..... ||||| .....
43813 TGGACAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 43862
   ..... ||||| .....
50 rGlyValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnArg 66
   ||||| ..... ||||| .....
43863 GGAAGAAGAAGAGG.....AAGTGGAGGAGCCGAGGAGGAGGAGGAGGAG 43897
   ..... ||||| .....
67 HisValAspArgAsnGluGlyGluAspValGlyArgPheValValGlu 83
   ..... ||||| .....
43898 GAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 43947
   ..... ||||| .....
83 nGlyThrGluValLysArgLysThrThrGluGlnGlnValArgProTyrA 100
   ||||| ..... ||||| .....
43948 AGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 43997
   ..... ||||| .....
100 rGArg 101
   .....
43998 GCAAA 44002
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH45143

seq_documentation_block:

ID AAH45143 standard; cDNA; 792 BP.

XX AAH45143;

DT 07-SEP-2001 (first entry)

XX Human brain expressed X-linked protein, hBex, coding sequence.

XX Human: brain expressed X-linked protein; cytostatic; auditory; nootropic;
KW hBex; dysembryoplasia; hereditary disease; cancer; tumour; deafness;
KW X-chromosome-binding mental retardation; lissencephalous disease; ss.

XX Homo sapiens.

OS WO200140286-A1.

PN 07-JUN-2001.

XX 27-NOV-2000; 2000WO-CN00502.

XX 30-NOV-1999; 99CN-0124179.

XX (BIOR-) BIORAD GENE DEV LTD SHANGHAI.

XX Mao Y, Xie Y;

XX WPI; 2001-397944/42.

XX P-PSDB; AAB99224.

XX Isolated human brain-expressed X-linked polypeptide used to diagnose
XX and treat of dysembryoplasia, hereditary diseases, cancer, tumor,
XX deafness and X-chromosome-binding mental retardation

XX Claim 5; Page 22; 30pp; Chinese.

score_list:	sequence	Strd Orig	ZScore	EScore	Len	Documentation
	/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-931-999-4 +		80.00	137.69	5.09	6755 !
	/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-936-907-179 -		74.50	103.29	420.15	56520 !
	/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-996-306-1 +		71.50	96.55	997.04	56516 !
	/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-938-907-1 -		71.50	96.55	997.04	56516 !
	/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-09-361-083-37 +		71.00	134.11	8.06	1360 !
	/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-09-258-373-21 +		71.00	130.78	12.36	1875 !
	/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-841-483-3 +		71.00	134.33	28.27	3490 !
	/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-382-911-3 +		71.00	134.33	28.27	3490 !
	/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-841-483-5 +		71.00	132.67	34.97	4094 !
	/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-382-911-5 +		71.00	132.67	34.97	4094 !
	/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-402-282-3 +		71.00	108.75	208.65	15664 !
	/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-508-004-3 +		71.00	108.75	208.65	15664 !
	/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-402-066-3 +		71.00	108.75	208.65	15664 !
	/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-402-068-3 +		71.00	108.75	208.65	15664 !
	/cgn2_6/ptodata/2/ina/PCRUS.COMB.seq:PCR-US93-03077-4 +		70.50	133.93	8.25	1242 !
	/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-198-4468-9 +		70.50	131.00	12.02	1647 !
	/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-870-693-9 +		70.50	131.00	12.02	1647 !
	/cgn2_6/ptodata/2/ina/PCRUS.COMB.seq:PCR-US93-03077-2 +		70.50	123.85	30.05	3279 !
	/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-105-537-34 +		70.50	120.14	48.38	4689 !
	/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-105-537-5 +		70.50	98.76	750.67	36778 !
	/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-09-320-878-19 +		70.50	98.29	797.98	38506 !
	/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-103-840A-2 +		70.00	47.97	3.6e+05	4403765 !
	/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-403-852D-2 +		69.50	131.47	11.32	1268 !
	/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-510-646B-2 +		69.50	131.47	11.32	1268 !
	/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-231-818-2 +		69.50	131.47	11.32	1268 !
	/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-403-852D-1 +		69.50	116.45	77.72	5392 !
	/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-510-646B-1 +		69.50	116.45	77.72	5392 !
	/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-231-818-1 +		69.50	116.45	77.72	5392 !
	/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-103-840A-2 -		69.50	46.85	4.0e+05	4403765 !
	/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-103-840A-1 -		69.50	46.83	4.0e+05	4411529 !
	/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-948-705-3 +		69.00	130.10	13.48	1298 !
	/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-09-120-365-89 +		69.00	127.07	19.89	1738 !
	/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-515-039-89 +		69.00	127.07	19.89	1738 !
	/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-09-120-365-4 +		69.00	126.49	21.44	1839 !
	/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-515-039-4 +		69.00	126.49	21.44	1839 !
	/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-826-246-1 +		69.00	125.86	23.23	1953 !
	/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-944-493-1 +		69.00	125.86	23.23	1953 !
	/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-09-126-640-1 +		69.00	125.86	23.23	1953 !
	/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-925-588-1 +		69.00	125.86	23.23	1953 !
	/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-07-807-043B-5 -		68.50	115.63	86.30	4698 !

```

Align seg 1/1 to: US-08-931-999-4 from: 1 to: 6755
14 GULYASPLSTYSASPYSARGLYGCLYLSALASerLysGlnSerGI 30
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
5974 GAAAAAACAGGAGGAAAGAAACAAAAGGAGGAGACGAGAAAAACAA 6023
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
30 uGLUGLuproHtHisLeuGLUGluValGLuAsnLysLysproGLYla 47
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```



```
LOCATION: 13617..13636
OTHER INFORMATION: downstream amplification primer 4-72, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13547..13564
OTHER INFORMATION: upstream amplification primer 4-73
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13962..13981
OTHER INFORMATION: downstream amplification primer 4-73, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 15994..16011
OTHER INFORMATION: downstream amplification primer 99-610
FEATURE:
NAME/KEY: primer_bind
LOCATION: 16463..16480
OTHER INFORMATION: upstream amplification primer 99-610, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 17304..17324
OTHER INFORMATION: downstream amplification primer 99-609
FEATURE:
NAME/KEY: primer_bind
LOCATION: 17814..17832
OTHER INFORMATION: upstream amplification primer 99-609, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18008..18027
OTHER INFORMATION: upstream amplification primer 4-90
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18423..18442
OTHER INFORMATION: downstream amplification primer 4-90, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18699..18716
OTHER INFORMATION: downstream amplification primer 99-607
FEATURE:
NAME/KEY: primer_bind
LOCATION: 19164..19182
OTHER INFORMATION: upstream amplification primer 99-607, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 22589..22609
OTHER INFORMATION: downstream amplification primer 99-602
FEATURE:
NAME/KEY: primer_bind
LOCATION: 23111..23129
OTHER INFORMATION: upstream amplification primer 99-602, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 25098..25118
OTHER INFORMATION: downstream amplification primer 99-600
FEATURE:
NAME/KEY: primer_bind
LOCATION: 25657..25674
OTHER INFORMATION: upstream amplification primer 99-600, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 26537..26557
OTHER INFORMATION: downstream amplification primer 99-598
FEATURE:
NAME/KEY: primer_bind
LOCATION: 27022..27040
OTHER INFORMATION: upstream amplification primer 99-598, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 32262..32281
OTHER INFORMATION: downstream amplification primer 99-592
FEATURE:
NAME/KEY: primer_bind
LOCATION: 32823..32841
```

```
OTHER INFORMATION: upstream amplification primer 99-592, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 34215..34233
OTHER INFORMATION: upstream amplification primer 99-217
FEATURE:
alignment_scores:
Quality: 74.50 Length: 82
Ratio: 1.656 Gaps: 2
Percent Similarity: 54.878 Percent Identity: 30.488
alignment_block:
US-09-327-750D-35 x US-09-338-907-179/rev ..
Align seg 1/1 to reverse of: US-09-338-907-179 from: 1 to: 56520
14 GtLysAspLysLysAspLysArgGlyGlyLysAlaSerLysGlnSerG1 30
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
39085 GAAAGCGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 39036
30 uGluCluProHisHisLeuGluGluValGluAsnLysLysProGlyGlyA 47
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
39035 AACGAGTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 38986
47 snValArgArgLysValArgArgLeuValProAsnPheLeuTrpAlaIle 63
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
38985 GAGAGAGAGCGAAAGAAAGAGG..... 38963
64 ProAsnArgHisValAspArgAsnGluGlyGlyGlu.....As 76
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
38962 ..... 38931
76 pValGlyArgPheValValGlnGlyThrGluValLysArgLysThr 91
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
38930 AATTGGGCATCTAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGTGT 38885
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-996-306-1
seq_documentation_block:
Sequence 1, Application US/08996306
Patent No. 5945522
GENERAL INFORMATION:
APPLICANT: Cohen, Daniel
APPLICANT: Chumakov, Ilya
APPLICANT: Blumenfeld, Marta
APPLICANT: Bougueleret, Lydie
TITLE OF INVENTION: prostate cancer gene
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSER: Knobbe, Martens, Olson & Bear
STREET: 501 West Broadway
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-3505
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Win95
SOFTWARE: Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/996,306
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: GENSET.018A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 1:
```

```

; NAME/KEY: SEQ ID65
; LOCATION: 51149..51168
; FEATURE:
; NAME/KEY: SEQ ID59
; LOCATION: 51448..51494
; FEATURE:
; NAME/KEY: SEQ ID68
; LOCATION: compl(51482..51499)
; FEATURE:
; NAME/KEY: SEQ ID44
; LOCATION: 51596..51613
; FEATURE:
; NAME/KEY: SEQ ID26
; LOCATION: 51612..51658
; FEATURE:
; NAME/KEY: SEQ ID53
; LOCATION: compl(51996..52015)
; FEATURE:
; NAME/KEY: polyad signal
; LOCATION: 54445..54450
;
US-08-996-306-1

alignment_scores:
    Quality: 71.50
    Ratio: 1.663
    Percent Similarity: 54.430
    Percent Identity: 30.380

alignment_block:
    US-09-327-750D-35 x US-08-996-306-1/rev ..

Align seg 1/1 to reverse of: US-08-996-306-1 from: 1 to: 56516

17 LysLysAspLysArgGlyGlyLysAlaSerLysGlnSerGluGluGluPr 33
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
39077 AAAAAAGAGAGAGAGAGAGAGAGCAAGCAAGCAAGAGAGAAACGAGTGA 39028
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
33 oHisHISLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 50
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
39027 GCGAAGGAAGAAAGAGGAGGAGGAGAAAGAAAGAAAGGAGGAGAGAG 38978
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
50 rGlyValAlaArgLeuValProAsnPheLeuTrpAlaIleProAsnArg 66
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
38977 CGAAAGAAAGAAAGG..... 38964

67 HisValAspArgAsnGluGlyGlyGlu..... AspValGlyArg 79
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
38963 .....AAGAAGAAAGGGGGAGAGCAACATGAGTTCAAATTGGCCA 38923

79 gpheValValGlnGlyThrGluValLysArgLysThr 91
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
38922 TACTGAGGCTCAAGGAACAAACATTC TAGAAGTAGT 38886

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-338-907-1

seq_documentation_block:
; Sequence 1, Application US/09338907
; Patent No. 6265546
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Cohen, Daniel
; APPLICANT: Ilya, Chumakov
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: PROSTATE CANCER GENE
; FILE REFERENCE: GENSET.18CPICP
; CURRENT APPLICATION NUMBER: US/09/338,907
; CURRENT FILING DATE: 1999-06-23
; EARLIER APPLICATION NUMBER: 08/996,306
; EARLIER FILING DATE: 1997-12-22
; EARLIER APPLICATION NUMBER: 60/099,658
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 09/218,207
; EARLIER FILING DATE: 1998-12-22

```



```

79 gpheValValGlnGlyThrGluValLysArgLysThr 91
:
:
:
38922 TACTGAGGCTGAAGCAACAAACATTCCTAGAGTACT 38886
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-961-083-37

seq_documentation_block:
; Sequence 37, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1360 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-083-37

alignment_scores:
Quality: 71.00 Length: 104
Ratio: 1.224 Gaps: 4
Percent Similarity: 55.769 Percent Identity: 25.962

alignment_block:
US-09-327-750D-35 x US-08-961-083-37
Align seg 1/1 to: US-08-961-083-37 from: 1 to: 1360

5 LysGlnValIleLeuAspLeuThrValGluLysAspLysArg 21
|||||:|||||:|||||:|||||:|||||:
413 AAAAAAGCGGAGCTTGAACATA...GTAAAGAGAGAACTAAGGAACCTCG 459

21 gGlyGlyLysAlaSerLysGlnSerGluGluGluProHisLysLeuGluG 38
|||||:|||||:|||||:|||||:|||||:
460 AACAGGAGAAAAGTTAGCAGCAANA.....GCGG 491

38 luValGluAsnLysLysProGlyGlyAsnValArgLysValArg 54
|||||:|||||:|||||:|||||:|||||:
492 AAGTTGAGAGTAAAAAAGCTCAGGCTACAGGTTAGAAAAATCAAG... 538

55 LeuValProAsnPheLeuTirPAlaIleProAsnArgHisValAspArg 71
|||||:|||||:|||||:|||||:|||||:

```



```

; SEQ ID NO 5
; LENGTH: 4094
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (126)..(3479)
US-08-841-483-5

```

```

alignment_scores:
  Quality: 71.00      Length: 82
  Ratio: 1.690        Gaps: 5
  Percent Similarity: 51.220  Percent Identity: 31.707

```

```
alignment_block:
```

```
US-09-327-750D-35 x US-08-841-483-5 ..
```

```
Align seg 1/1 to: US-08-841-483-5 from: 1 to: 4094
```

```

12 ThrValGluLysAspLysLysAspLysArgGlyGly.....LysAL 25
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1452 ACTCTGAACGAACGAAGAGAGAGAGAGAGGCGATCTTCAAGAGAGAGTC 1501
25 aserLysGlnSerGluGluGlu..... 32
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1502 CAGCAAGAAAGGCGCTGAGGAGGCGCGCTGGAGACCCCTTCATCATCAGGC 1551
33 .....ProHisHisLeuGluGlu.....ValGluAsnLysLys 43
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1552 CCACCCCGCTCCCGCTCATGAAGCCCTGCTGGTGTGTTGTGAACCCCAAG 1601
44 ProGlyGlyAsnValArgArgLysValArgLeuValProAsnPhel 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1602 AGTGGGGCAACGAGGTGCAGATC.....ATCCAGTCTTCT 1642
60 utrPalalleProAsnArgHisVal....AspArgAsnGluGlyGly 74
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1643 CTGGTATCTCAATCCCGACAGTCTTCGACCTGAGCCAGGAGGG 1688
seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-382-911-5

```

```
seq_documentation_block:
```

```

; Sequence 5, Application US/09382911
; Patent No. 6221658
; GENERAL INFORMATION:
; APPLICANT: Prescott, Steven M.
; APPLICANT: Bunting, Michaeline
; APPLICANT: Tang, Wen
; APPLICANT: Topham, Matthew
; TITLE OF INVENTION: Diacylglycerol Kinase Isoforms Epsilon and Zeta and
; FILE REFERENCE: 2037.2.1a
; CURRENT APPLICATION NUMBER: US/09/382,911
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: 08/841,483
; PRIOR FILING DATE: 1997-04-22
; PRIOR APPLICATION NUMBER: 60/016,210
; PRIOR FILING DATE: 1996-04-22
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 4094
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (126)..(3479)
US-09-382-911-5

```

```
alignment_scores:
```

```
Quality: 71.00      Length: 82
```

```

Ratio: 1.690        Gaps: 5
Percent Similarity: 51.220  Percent Identity: 31.707
alignment_block:
US-09-327-750D-35 x US-09-382-911-5 ..
Align seg 1/1 to: US-09-382-911-5 from: 1 to: 4094
12 ThrValGluLysAspLysLysAspLysArgGlyGly.....LysAL 25
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1452 ACTCTGAACGAACGAAGAGAGAGAGGCGATCTTCAAGAGAGAGTC 1501
25 aserLysGlnSerGluGluGlu..... 32
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1502 CAGCAAGAAAGGCGCTGAGGAGGCGCGCTGGAGACCCCTTCATCATCAGGC 1551
33 .....ProHisHisLeuGluGlu.....ValGluAsnLysLys 43
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1552 CCACCCCGCTCCCGCTCATGAAGCCCTGCTGGTGTGTTGTGAACCCCAAG 1601
44 ProGlyGlyAsnValArgArgLysValArgLeuValProAsnPhel 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1602 AGTGGGGCAACGAGGTGCAGATC.....ATCCAGTCTTCT 1642
60 utrPalalleProAsnArgHisVal....AspArgAsnGluGlyGly 74
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1643 CTGGTATCTCAATCCCGACAGTCTTCGACCTGAGCCAGGAGGG 1688
seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-402-282-3

```

```
seq_documentation_block:
```

```

; Sequence 3, Application US/08402282
; Patent No. 5476768
; GENERAL INFORMATION:
; APPLICANT: Pearson, Robert E.
; APPLICANT: Dickson, Julie A.
; APPLICANT: Hamilton, Paul T.
; APPLICANT: Little, Michael C.
; APPLICANT: Beyer Jr., Wayne E.
; TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
; TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
; ADDRESSEE: Company
; STREET: 1 Becton Drive
; CITY: Franklin Lakes
; STATE: NJ
; COUNTRY: US
; ZIP: 07417
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/402,282
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fugit, Donna R.
; REGISTRATION NUMBER: 32,135
; REFERENCE/DOCKET NUMBER: P-3283
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15664 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 222..425

```

```

; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 451..747
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 747..1109
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1109..2014
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2034..2747
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2747..3109
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3109..3444
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3444..3728
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3731..4855
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4855..5376
; OTHER INFORMATION: /function- "potential coding
; OTHER INFORMATION: sequence"
; OTHER INFORMATION: /product- "L5 gp37 homolog"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5382..5747
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5837..6307
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6403..7770
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7770..8006
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8033..8236
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"

```

```

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8244..9443
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9450..10244
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10371..10586
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11115..11786
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11917..12741
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 12748..14499
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14771..15154
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15154..15426
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15429..15664
; OTHER INFORMATION: /function- "potential open reading
; OTHER INFORMATION: frame"
; US-08-402-282-3
; alignment_scores:
; Quality: 71.00 Length: 71
; Ratio: 1.775 Gaps: 4
; Percent Similarity: 56.338 Percent Identity: 33.803
; alignment_block:
; US-09-327-750D-35 x US-08-402-282-3
; Align seg 1/1 to: US-08-402-282-3 from: 1 to: 15664
; 49 ArgArgLysValArgLeuValProAsnPheLeuTrpAlaIleProAs 65
; 13436 CGAAATCGGTGGCGCAAGGCGCGCGATGCGCGTGG.....CGAC 13479
; 65 narg.His.....ValAspArgAsnGluGly 73
; 13480 GCGCTCACCCCGAACCGGACCTCGAGTCCGGTCGATCGGTCCGGGA 13529
; 74 GlyGluAspValGlyArgPheValGlnGlyThrGluValLysArgLy 90
; 13530 GGGATGAT.....CGAAGCTACCGCGCGGACACGAGAAAACGGCG 13573
; 90 sThrThrGluGlnValArgProTyrArgPheArgThrProGluP 107
; 13574 GCGACGACGAGGAATCGACGCGCCGACGACGACGATCGTGACGTC 13623

```

```
107 roaspsnHls 110
|||:|||||
13624 CTTTCAGCAT 13634

seq_name: /cgn2_6/ptodata/2/ins/5A_COMB.seq:US-08-508-004-3

seq_documentation_block:
: Sequence 3, Application US/08508004
: Patent No. 5582969
: GENERAL INFORMATION:
: APPLICANT: Pearson, Robert E.
: APPLICANT: Dickson, Julie A.
: APPLICANT: Hamilton, Paul T.
: APPLICANT: Little, Michael C.
: APPLICANT: Beyer Jr., Wayne F.
: TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
: TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
: ADDRESSEE: Company
: STREET: 1 Becton Drive
: CITY: Franklin Lakes
: STATE: NJ
: COUNTRY: US
: ZIP: 07417
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/508,004
: FILING DATE: 27-JUL-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/402,282
: FILING DATE: 10-MAR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Fugit, Donna R.
: REGISTRATION NUMBER: 32,135
: REFERENCE/DOCKET NUMBER: P-3283
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 15864 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 222..425
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 451..747
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 747..1109
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 1109..2014
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 2034..2747
```

```
:
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 2747..3109
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 3109..3444
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 3444..3728
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 3731..4855
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 4855..5376
: OTHER INFORMATION: /function= "potential coding
: OTHER INFORMATION: sequence"
: OTHER INFORMATION: /product= "L5 gp37 homolog"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 5382..5747
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 5837..6307
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 6403..7770
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 7770..8006
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 8033..8236
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 8244..9443
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 9450..10244
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 10371..10586
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 11115..11786
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
```



```

FEATURE:
NAME/KEY: misc.feature
LOCATION: 11917..12741
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 12748..14499
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 14771..15134
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 15154..15426
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 15429..15664
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
US-08-508-004-3
alignment_scores:
Quality: 71.00 Length: 71
Ratio: 1.775 Gaps: 4
Percent Similarity: 56.338 Percent Identity: 33.803
alignment_block:
US-09-327-750D-35 x US-08-508-004-3
Align seg 1/1 to: US-08-508-004-3 from: 1 to: 15664
49 ArgArgLysValArgAtrGLeuValProAspPheLeuTrpAlaIleProAs 65
|||||
13436 CGAATCGGTGGCGGGAAGGGCGCGGATGCCGTGG.....CCGAC 13479
|||||
65 Arg.His.....ValAspArgAsnGluGly 73
|||||
13480 CGGCTCACCCGGAACCCGACCTCGAGTCCGTCGTCGTCGTCGCGGA 13529
|||||
74 GlyGluAspValGlyArgPheValValGlnGlyThrGluValLysArgly 90
|||||
13530 GGGATGAT.....CGAGCCTACCGCGGACACGAGAGAAACGGCG 13573
|||||
90 sThrThrGluGlnValArgProTyrArgArgPheArgThrProGluP 107
|||||
13574 GCGAGCGAGCGGAATCGACGGCCGACGACGACGATCGTCGTCGATC 13623
|||||
107 roAspAsnHis 110
||
13624 CTGTTCCAGCAT 13634
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-402-066-3
seq_documentation_block:
Sequence 3, Application US/08402066
Patent No. 5612182
GENERAL INFORMATION:
APPLICANT: Pearson, Robert E.
APPLICANT: Dickson, Julie A.
APPLICANT: Hamilton, Paul T.
APPLICANT: Little, Michael C.
APPLICANT: Beyer Jr., Wayne F.
TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:

```

```

ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
ADDRESSEE: Company
STREET: 1 Becton Drive
CITY: Franklin Lakes
STATE: NJ
COUNTRY: US
ZIP: 07417
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/402,066
FILING DATE:
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Fugit, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-3283
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 15664 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc.feature
LOCATION: 222..425
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 451..747
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 747..1109
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1109..2014
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 2034..2747
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 2747..3109
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 3109..3444
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 3444..3728
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 3731..4855
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:

```

```

; NAME/KEY: misc_feature
; LOCATION: 4855..5376
; OTHER INFORMATION: /function= "potential coding
; OTHER INFORMATION: sequence"
; OTHER INFORMATION: /product= "L5 gp37 homolog"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5382..5747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5837..6307
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6403..7770
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7770..8006
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8033..8236
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8244..9443
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9450..10244
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10371..10586
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11115..11786
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11917..12741
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 12748..14499
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14771..15154
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15154..15426
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15429..15664

```

```

; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; US-08-402-066-3

alignment_scores:
  Quality: 71.00      Length: 71
  Ratio: 1.775       Gaps: 4
  Percent Similarity: 56.338  Percent Identity: 33.803

alignment_block:
  US-09-327-750D-35 x US-08-402-066-3

Align seg 1/1 to, US-08-402-066-3 from: 1 to: 15664
  49 ArgArgLysValArgLeuValProAsnPheLeuTrpAlaIleProAs 65
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  13436 CGAATCGGTGGCGCAAGGCGCGCGATGGCGGTGG.....CCGAC 13479
  65 nArg.His'.....ValasPArgAsnGluGly 73
  |||:|||||:|||||:|||||:|||||:|||||:|||||:
  13480 GCGCTCACCCCGAACCGGACCTCGAGGTCCGGTCGATCGGTCCGGGA 13529
  74 GlyGluAspValGlyArgPheValValGlnGlyThrGluValLysArgLy 90
  |||:|||||:|||||:|||||:|||||:|||||:|||||:
  13530 GGGATGAT.....CGAAGCCTACCGGCGGACACGAAGAAACGGCG 13573
  90 sThrThrGluGlnValArgProTyrArgArgPheArgThrProGluP 107
  :|||||:|||||:|||||:|||||:|||||:|||||:
  13574 GCGACGACGACGGAATCGACGGCCGACGACGACATCGTCGATCGTCGATC 13623
  107 roAsPAsnHis 110
  |||:|||||:
  13624 CTGTTCCAGCAT 13634

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-402-068-3

seq_documentation_block:
; Sequence 3, Application US/08402068
; Patent No. 5633159
; GENERAL INFORMATION:
; APPLICANT: Pearson, Robert E.
; APPLICANT: Dickinson, Julie A.
; APPLICANT: Hamilton, Paul T.
; APPLICANT: Little, Michael C.
; APPLICANT: Beyer Jr., Wayne F.
; TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
; ADDRESSEE: Company
; STREET: 1 Becton Drive
; CITY: Franklin Lakes
; STATE: NJ
; COUNTRY: US
; ZIP: 07417
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/402,068
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fugit, Donna R.
; REGISTRATION NUMBER: 32,135
; REFERENCE/DOCKET NUMBER: P-3283
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15664 base pairs

```


OM of: US-09-327-750D-35 to: GenEmbl.* out_format : pfs
 Date: Mar 11, 2002 3:35 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framet_p2n_model -DEV=xlh
 -O=/cgn2_1/USFTO_spool/US09327750/runat_11032002_101154_20319/app_query.fasta_1.1472
 -DB=GenEmbl -QPM=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -DELAEXT=7.000 -XGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blossum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTPMT=pfs -NORM=ext -HEAPSIZE=500
 -MINLEN=0 -MAXLEN=2000000000 -USER=US09327750.ecgn1_1_8673
 -NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-327-750D-35
 Query length: 117
 Database: GenEmbl.*
 Database sequences: 1472140
 Database length: -341344837
 Search time (sec): 4557.230000

score list:

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
gb_pr:AK000959	+	355.50	626.28	1.3e-26	1229	AK000959 Homo sapiens cDNA FLJ
gb_pr:AI00231	+	355.50	625.44	1.5e-26	1364	AI00231 Sequence 32 from Pat
gb_pr:HS635G19	+	355.50	593.78	8.5e-25	69648	AI0035494 Human DNA sequence f
gb_pr:BC003190	+	175.00	306.91	8.1e-09	793	BC003190 Homo sapiens, p75NTR-a
gb_pr:HUMOGC	+	175.00	305.97	9.1e-09	891	M38188 Human unknown protein fr
gb_pr:AF187064	+	175.00	305.97	9.1e-09	891	AF187064 Homo sapiens p75NTR-as
gb_pr:AF187066	+	154.00	270.34	8.8e-07	700	AF187066 Mus musculus p75NTR-as
gb_ro:AF097440	-	154.00	268.74	1.1e-06	854	AF097440 Mus musculus brain exp
gb_sts:G27208	-	150.00	267.28	1.3e-06	421	G27208 MARC 4953-4954:99199031
gb_ro:AF187065	+	149.50	264.70	1.8e-06	519	AF187065 Rattus norvegicus p75N
gb_pr:HSW351F8	+	138.00	208.08	0.0026	45678	G270719 Human DNA sequence fr
gb_sts:G35294	-	120.50	213.50	0.0013	477	G35294 Human STS SHGC-37409, se
gb_pr:HS71487	+	114.50	159.87	1.25	98274	I299755 Human DNA sequence fr
gb_ro:AF097438	+	107.50	185.74	0.0454	835	AF097438 Mus musculus brain exp
gb_ro:AF097437	+	107.50	185.71	0.0456	838	AF097437 Mus musculus REX-3 mR
gb_ro:AF097439	+	107.50	176.69	0.1275	2269	AF097439 Mus musculus brain exp
gb_pr:AF251053	+	99.50	172.35	0.1506	785	AF251053 Homo sapiens X-linked
gb_pr:AF251052	+	99.50	170.84	0.3070	898	AF251052 Sequence 76 from Paten
gb_pr:HSV870H8	+	99.50	142.25	12.02	31321	I270233 Human DNA sequence fr
gb_pr:AL133348	+	99.50	140.16	15.70	40584	AL133348 Human DNA sequence fr
gb_htg:HSU80B1	+	99.50	140.07	15.88	41029	AL022169 Homo sapiens chromos
gb_pr:HS75399	+	98.50	129.88	58.71	116561	AL023653 Human DNA sequence
gb_pr:AF195953	+	98.50	125.19	107.09	208618	AF195953 Homo sapiens membra
gb_htg:AC012947	+	95.00	121.94	162.57	143599	AC012947 Drosophila melanoga
gb_in:AC009183	+	95.00	119.60	219.49	192055	AC009183 Drosophila melanoga
gb_in:AE003685	+	95.00	118.32	221.08	193398	AE003685 Drosophila melanoga
gb_pr:AF183416	+	94.00	163.70	0.7668	691	AF183416 Homo sapiens ovarian g
gb_pr:AF237783	+	94.00	162.02	0.9513	742	AF237783 Homo sapiens brain-exp
gb_pr:AF220189	+	94.00	161.65	0.9972	828	AF220189 Homo sapiens uncharact
gb_pr:HS19894	+	94.00	131.35	48.61	35714	AL008708 Human DNA sequence f
gb_pr:AC002478	+	91.00	120.55	194.16	70113	AC002478 Human BAC clone GSI
gb_htg:AC012232	+	91.00	112.22	497.28	174337	AC012232 Homo sapiens clone
gb_da:WAFR2	+	90.50	146.47	6.99	2507	J04157 M.xanthus fr2A, fr2B an
gb_pr:AP000265	+	90.50	124.52	116.65	38304	AP000265 Homo sapiens genomic
gb_pr:AP000034	+	90.50	116.80	314.19	100000	AP000034 Homo sapiens genomic
gb_pr:AP000101	+	90.50	116.80	314.19	100000	AP000101 Homo sapiens genomic
gb_pr:AP000177	+	90.50	116.80	314.19	100000	AP000177 Homo sapiens genomic
gb_htg:AC026930	+	90.50	114.66	413.39	130443	AC026930 Homo sapiens chromo

gb_pr:AC005295 + 90.50 114.62 415.44 131070 ! AC005295 Homo sapiens xp2
 gb_pr:AP001713 + 90.50 106.95 1.1e+03 340000 ! AP001713 Homo sapiens gen
 gb_da:AE008371 - 90.00 133.00 39.36 11967 ! AE008371 Agrobacterium tum
 gb_pr:U72787 + 90.00 129.48 61.75 18510 ! U72787 Homo sapiens cosmid
 gb_pr:AL353806 + 90.00 115.06 392.66 111051 ! AL353806 Human DNA sequen

seq_name: gb_pr:AK000959

seq documentation block:

LOCUS AK000959 1229 bp mRNA PRI 22-FEB-2000
 DEFINITION Homo sapiens cDNA FLJ10097 fls, clone HEMBA1002458, weakly similar
 to OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74.
 ACCESSION AK000959
 VERSION AK000959.1 GI:7021945
 KEYWORDS oligo kapping; fls (full insert sequence).
 SOURCE Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to
 mRNA, clone lib:HEMBA1 clone:HEMBA1002458.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (sites)
 AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
 Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H.,
 Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K.,
 Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
 Wakamatsu,A., Nakamura,Y., Negahari,K., Masuno,Y. and Oshima,A.
 NEDO human cDNA sequencing project
 UNPUBLISHED (2000)
 REFERENCE 2 (bases 1 to 1229)
 AUTHORS Isogai,T. and Otsuki,T.
 TITLE Direct Submission
 JOURNAL Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao
 Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana,
 Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
 Tel:81-438-52-3951, Fax:81-438-52-3952)

COMMENT

NEDO human cDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan; cDNA full insert
 sequencing: Research Association for Biotechnology; cDNA library
 construction, 5'- and 3'-end one pass sequencing and clone selection:
 Helix Research Institute (supported by Japan Key Technology Center
 etc.) and Department of Virology, Institute of Medical Science,
 University of Tokyo.

FEATURES

Location/Qualifiers
 1..1229
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HEMBA1002458"
 /clone_lib="HEMBA1"
 /dev_stage="embryo, 10 weeks"
 /tissue_type="whole embryo, mainly head"
 /note="cloning vector: pME18SFL3"
 191..553
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAA91443.1"
 /db_xref="GI:7021946"

CDS

translation="MESKEELANLNNGENAGQENGGEOAPTONEEESRHLGGGEGO
 KPGCNIRGVRRLVPRVPRWAIPIRHLIEHNEARDVPRFGOMMEIKRKRTRQQMRHY
 MRFQPEPDNHYDFCLIP"
 351 a 243 c 296 g 339 t

BASE COUNT

ORIGIN

alignment_scores:
 Quality: 355.50 Length: 120
 Ratio: 3.665 Gaps: 3
 Percent Similarity: 80.833 Percent Identity: 59.167

alignment_block:

US-09-327-750D-35 x AK000959
 Align seg 1/1 to: AK000959 from: 1 to: 1229

```

1 MetAlaSerLysLysGlnValIleLeu...AspLeuThrValGluLysAs 16
111 |||||.....:|||||.....:|||||.....:|||||.....:
191 ATGGAGTCCAAAGAGGAAGTACGGCAACAATCTCAACGGGGAATATGC 240
16 physLysAspLysArgGlyGly...LysAlaSerLysGlnSerGluGlu 32
111 |||||.....:|||||.....:|||||.....:|||||.....:
241 CCAACAAGAAACCAAGAGGGGAGGAGCCGCCAGCAAGATGAAGAAG 290
32 LuProHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnVal 48
111 |||||.....:|||||.....:|||||.....:|||||.....:
291 AATCCCGCCATTTGGGAGGGGGTGAAGCCAGAGCCTGGAGGAATATC 340
49 ArgArg...LysValArgArgLeuValProAsnPheLeuTrpAlaIlePr 64
111 |||||.....:|||||.....:|||||.....:|||||.....:
341 AGCGGGGGCGAGTTAGCGCACTGTCCCTAATTTTCGATGGCCATACC 390
64 QAsnArgHisValAspArgAsnGluGlyGlyGluAspValGlyArgPheV 81
111 |||||.....:|||||.....:|||||.....:|||||.....:
391 TAATAGGCATATTGAGCACAAATGAAGCGAGAGATGATGATGAAGAAGTTTG 440
81 alValGlnGlyThrGluValLysArgLysThrThrGluGlnGlnValArg 97
111 |||||.....:|||||.....:|||||.....:|||||.....:
441 TAGGCACAGATGATGCAATCAAGAAAGACTAGGGAACAGCAGATGAGG 490
98 ProTyArgArgPheArgThrProGluProAspAsnHisTyArgPheCy 114
111 |||||.....:|||||.....:|||||.....:|||||.....:
491 CACTATATGCGCTTCCAAACTCTGAACTGACACCACTATGACTTTTG 540
114 sleuilePro 117
111 |||||.....:|||||.....:|||||.....:|||||.....:
541 CCTCATACCT 550

```

seq_name: gb_pat:AX100231

seq_documentation_block:

LOCUS AX100231 1364 bp DNA PAT 02-APR-2001
 DEFINITION Sequence 32 from Patent WO0119860.
 ACCESSION AX100231
 VERSION AX100231.1 GI:19539115
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1364)
 AUTHORS Tang,F.T., Hillman,J.L., Yue,H., Reddy,R., Lal,P., Shah,P.,
 Azimzai,Y., Baughn,M.R., Lu,D.A., Bandman,O., Shih,L.L. and
 Patterson,C.
 TITLE Proteins associated with cell differentiation
 JOURNAL Patent: WO 0119860-A 32 22-MAR-2001;
 Incyte Genomics, Inc. (US)
 FEATURES
 source
 1..1364
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="Incyte ID No: 1990956CB1"

BASE COUNT 411 a 269 c 322 g 362 t
 ORIGIN
 alignment_scores:
 Quality: 355.50 Length: 120
 Ratio: 3.665 Gaps: 3
 Percent Similarity: 80.833 Percent Identity: 59.167

alignment_block:

US-09-327-750D-35 x AX100231

Align seg 1/1 to: AX100231 from: 1 to: 1364

```

1 MetAlaSerLysLysGlnValIleLeu...AspLeuThrValGluLysAs 16
111 |||||.....:|||||.....:|||||.....:|||||.....:
250 ATGGAGTCCAAAGAGGAAGTACGGCAACAATCTCAACGGGGAATATGC 299

```

```

16 physLysAspLysArgGlyGly...LysAlaSerLysGlnSerGluGlu 32
111 |||||.....:|||||.....:|||||.....:|||||.....:
300 CCAACAAGAAACCAAGAGGGGAGGAGCCGCCAGCAGATGAAGAAG 349
32 LuProHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnVal 48
111 |||||.....:|||||.....:|||||.....:|||||.....:
350 AATCCCGCCATTTGGGAGGGGGTGAAGCCAGAGCCTGGAGGAATATC 399
49 ArgArg...LysValArgArgLeuValProAsnPheLeuTrpAlaIlePr 64
111 |||||.....:|||||.....:|||||.....:|||||.....:
400 AGCGGGGGCGAGTTAGCGCACTGTCCCTAATTTTCGATGGCCATACC 449
64 QAsnArgHisValAspArgAsnGluGlyGlyGluAspValGlyArgPheV 81
111 |||||.....:|||||.....:|||||.....:|||||.....:
450 TAATAGGCATATTGAGCACAAATGAAGCGAGAGATGATGATGAAGAAGTTTG 499
81 alValGlnGlyThrGluValLysArgLysThrThrGluGlnGlnValArg 97
111 |||||.....:|||||.....:|||||.....:|||||.....:
500 TAGGCACAGATGATGCAATCAAGAAAGACTAGGGAACAGCAGATGAGG 549
98 ProTyArgArgPheArgThrProGluProAspAsnHisTyArgPheCy 114
111 |||||.....:|||||.....:|||||.....:|||||.....:
550 CACTATATGCGCTTCCAAACTCTGAACTGACACCACTATGACTTTTG 599
114 sleuilePro 117
111 |||||.....:|||||.....:|||||.....:|||||.....:
600 CCTCATACCT 609

```

seq_name: gb_pr:HS635G19

seq_documentation_block:

LOCUS HS635G19 69548 bp DNA PRI 23-NOV-1999
 DEFINITION Human DNA sequence from clone 635G19 on chromosome Xq22.1-22.3
 Contains a LAMR1 (Laminin Receptor 1 (67kD) (RPSA, 40S Ribosomal
 Protein SA, P40)) pseudogene and part of a novel protein. Contains
 ESTs and GSSs, complete sequence.
 ACCESSION AL035494
 VERSION HTG; 40S Ribosomal Protein P40; 40S Ribosomal Protein SA; Cpg
 KEYWORDS Island; Laminin Receptor 1; LAMR1; RPSA.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 69548)
 AUTHORS Bird,C.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUN-1999) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 COMMENT
 On May 11, 1999 this sequence version replaced gi:4678749.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL
 This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated human repeat sequence elements (e.g. Alu). Where the
 sequence is ambiguous, there is an annotation using the 'unsure'
 feature key.
 This sequence was generated from part of bacterial clone contigs of
 human chromosome X, constructed by the Sanger Centre Chromosome X
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/ChrX
 635G19 is from the library RPI14 constructed at the Roswell Park
 Cancer Institute by the group of Pieter de Jong. For further

details see <http://bacpac.med.buffalo.edu/VECTOR:pcvPAC2>
 IMPORTANT: This sequence is not the entire insert of clone 635G19.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we arrange for a small overlap between
 neighbouring submissions.
 The true left end of clone U65A4 (281014) is at 69549 in this
 sequence. The true right end of clone U101D3 (285997) is at 100 in
 this sequence.

FEATURES

Location/Qualifiers

source

1..69648
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="q22.1-22.3"
 /clone="RP4-635G19"
 /clone.lib="RPC1-4"
 439..483

repeat_region

/note="TAR1 repeat: matches 707..749 of consensus"
 501..561
 /note="AluJb repeat: matches 137..311 of consensus"
 662..897
 /note="118 copies 2 mer aa 70% conserved"
 915..1331
 /note="MSTA repeat: matches 1..419 of consensus"
 1334..1517
 /note="92 copies 2 mer ga 77% conserved"
 1927..2020
 /note="MIR repeat: matches 109..217 of consensus"
 2066..2262
 /note="L1ME2 repeat: matches 5825..6022 of consensus"
 2263..2574
 /note="AluSp repeat: matches 1..313 of consensus"
 2575..2819
 /note="L1ME2 repeat: matches 5583..5825 of consensus"
 2821..3039
 /note="Alu repeat: matches 1..311 of consensus"
 3104..3384
 /note="L1ME2 repeat: matches 5304..5591 of consensus"
 3391..4007
 /note="L1MB1 repeat: matches 5558..6162 of consensus"
 4012..4137
 /note="Tigger2a repeat: matches 11..123 of consensus"
 4141..4252
 /note="L1 repeat: matches 2920..3031 of consensus"
 4265..4350
 /note="HY4 repeat: matches 1..87 of consensus"
 4350..4678
 /note="MER74A repeat: matches 28..369 of consensus"
 4914..5095
 /note="MER5A repeat: matches 8..189 of consensus"
 5224..5555
 /note="L2 repeat: matches 2423..2746 of consensus"
 6040..6086
 /note="L2 repeat: matches 1676..1720 of consensus"
 7413..7478
 /note="Alu repeat: matches 239..304 of consensus"
 7628..7928
 /note="AluB8 repeat: matches 1..316 of consensus"
 7952..7987
 /note="18 copies 2 mer aa 81% conserved"
 7990..8305
 /note="AluJb repeat: matches 1..309 of consensus"
 8436..8538
 /note="MIR repeat: matches 99..212 of consensus"
 10836..11442
 /note="L1MD repeat: matches 1357..1972 of consensus"
 11443..11571
 /note="FLAM-A repeat: matches 2..129 of consensus"
 11572..11946
 /note="L1MD repeat: matches 977..1357 of consensus"
 11947..12245
 /note="Alu repeat: matches 1..301 of consensus"
 12246..13098

/note="L1MD repeat: matches 275..977 of consensus"
 13094..13223
 /note="L1MD repeat: matches 17..113 of consensus"
 13523..13820
 /note="MLTID repeat: matches 177..505 of consensus"
 13823..13932
 /note="55 copies 2 mer ct 77% conserved"
 13936..13977
 /note="MLTID repeat: matches 201..241 of consensus"
 13978..14240
 /note="AluSg1 repeat: matches 1..270 of consensus"
 14245..14534
 /note="AluSc repeat: matches 1..290 of consensus"
 14535..14721
 /note="MLTID repeat: matches 2..201 of consensus"
 14724..15604
 /note="L1MC/D repeat: matches 4646..5565 of consensus"
 15605..15909
 /note="AluX repeat: matches 24..312 of consensus"
 16069..16363
 /note="AluX repeat: matches 1..291 of consensus"
 16482..16641
 /note="AluJb repeat: matches 136..295 of consensus"
 16671..16953
 /note="AluX repeat: matches 1..284 of consensus"
 17184..17280
 /note="L1P repeat: matches 4499..4619 of consensus"
 17326..17377
 /note="26 copies 2 mer at 79% conserved"
 17704..17863
 /note="AluSc repeat: matches 118..295 of consensus"
 17864..18175
 /note="AluB8 repeat: matches 1..318 of consensus"
 18176..18298
 /note="AluSc repeat: matches 1..118 of consensus"
 18848..18873
 /note="113 copies 2 mer tg 100% conserved"
 18876..19171
 /note="AluSg repeat: matches 1..295 of consensus"
 19183..19232
 /note="25 copies 2 mer ac 100% conserved"
 19246..19533
 /note="Alu repeat: matches 5..292 of consensus"
 19539..19837
 /note="MER74B repeat: matches 61..368 of consensus"
 20016..20194
 /note="MER74A repeat: matches 295..487 of consensus"
 20394..20692
 /note="AluSg repeat: matches 1..302 of consensus"
 20701..20809
 /note="LTR1 repeat: matches 682..785 of consensus"
 20877..21061
 /note="LTR1 repeat: matches 1..195 of consensus"
 21186..21636
 /note="LTR29 repeat: matches 3..498 of consensus"
 21609..21737
 /note="MER34 repeat: matches 409..543 of consensus"
 21771..22076
 /note="Alu repeat: matches 2..308 of consensus"
 22878..22974
 /note="MIR repeat: matches 92..191 of consensus"
 23452..23582
 /note="FLAM_C repeat: matches 1..131 of consensus"
 23583..23624
 /note="21 copies 2 mer ta 76% conserved"
 23735..24041
 /note="AluSg1 repeat: matches 1..308 of consensus"
 24045..24072
 /note="14 copies 2 mer tt 89% conserved"
 25274..25582
 /note="AluB repeat: matches 1..289 of consensus"
 25583..25747
 /note="match: GSS B63622"

misc_feature


```

|||||.....|
316 CGCCGACTTGGCCCTAATTTTCATGGGCCATACCAATAGGCAGATCAA 365
69 pArgAsnGluGly.....GlyGluAspValGlyArgPheValValGlnG 84
   |||||.....|
366 TGATGGGATGGGTGGAGATGGAGATGATGGAATATTCATGGAGGAGA 415
84 lyThrGluValLysArgLysThrThrGluGlnGlnValArgProTyrArg 100
   |||||.....|
416 TGAGAGAAATCAGAGAAATAGGAGCTGCAGTTGAGGAATTGTCTG 465
101 ArgPheArgThrProGluProAsnHisTyrAsp.....Ph 113
   |||||.....|
466 CGTATCCTTATGGGAGGCTCTTAATCACCATGACCATGATGATGAATT 515
113 eCysLeuIlePro 117
   |||||.....|
516 TTGCCTTATGCCT 528

```

seq_name: gb_pr:HUMOGC

```

seq_documentation_block:
LOCUS HUMOGC 891 bp mRNA 07-MAR-1995
DEFINITION Human unknown protein from clone pHR74 mRNA, complete cds.
ACCESSION M38188 X56942
VERSION M38188.1 GI:189378
KEYWORDS Human ovarian granulosa cell line, cDNA to mRNA.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 891)
AUTHORS Rapp, G., Freudenstein, J., Klaudiny, J., Mucha, J., Wempe, F.,
Zimmer, M. and Scheit, K.H.
TITLE Characterization of three abundant mRNAs from human ovarian
granulosa cells
JOURNAL DNA Cell Biol. 9 (7), 479-485 (1990)
MEDLINE 91025550
COMMENT Draft entry and computer-readable sequence for [DNA 9, 479-485
(1990)] kindly submitted
by K.H.Scheit, 27-AUG-1990.

```

FEATURES

```

source
1..891
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="pHR74"
/cell_type="granulosa"
/tissue_type="ovary"
<1..891
/note="protein of unknown function"
312..647
/note="protein of unknown function"
/codon_start=1
/protein_id="AAA63232.1"
/db_xref="GI:189379"
/translation="MANIHQENEFQPMONGEDRPLGGEGHQPAGNRNRGQARRLA
PNFHWAPNRIQDNGMGDDMEIFWEEMREIRKLRELQLRNLRLMGLSELNHH
HHDEFCLMP"
361..534
/note="protein of unknown function"
/codon_start=1
/protein_id="AAA63233.1"
/db_xref="GI:189380"
/translation="MERKTALEEVKATSLQIDGDLADLPFLFDGYPPIGRSMWGH
VEMIMKYSWRR"

```

CDS

```

BASE COUNT 251 a 182 c 224 g 234 t
ORIGIN

```

alignment_scores:

```

Quality: 175.00 Length: 121
Ratio: 2.273 Gaps: 4
Percent Similarity: 63.636 Percent Identity: 36.364

```

alignment_block:

US-09-327-750D-35 x HUMOGC

Align seg 1/1 to: HUMOGC from: 1 to: 891

```

4 LysLysGlnValIleLeu...AspLeuThrValGluLysAspLysLysAs 19
   |||||.....|
297 AAAAAAATCTCATCTGCAAAATATTCACCAAGAAACGAAGAGATGGA 346
19 pLysArgGlyGlyLysAlaSerLysGlnSerGluGluGluProHisHisL 36
   |||||.....|
347 GCAG.....CCTATGCAGAAATGGAGAGAGAGACCCGCTT 381
36 euGluGluValGluAsnLysLysProGlyGlyAsnValArgLysVal 52
   |||||.....|
382 TGGAGGAGGTGAAGGCCACCGCTGCAGAAATCGACGGGACAGGCT 431
53 ArgArgLeuValProAsnPheLeuTTPAlaIleProAsnArgHisValas 69
   |||||.....|
432 CGCCGACTTCCCTCAATTTTCATGGCCCATACCAATAGGCAGATCAA 481
69 pArgAsnGluGly.....GlyGluAspValGlyArgPheValValGlnG 84
   |||||.....|
482 TGATGGGATGGGTGGAGATGGAGATGATATGGAATATTCATGGAGGAGA 531
84 lyThrGluValLysArgLysThrThrGluGlnGlnValArgProTyrArg 100
   |||||.....|
532 TGAGAGAAATCAGAGAAACTTAGGGAGCTGCAGTTGAGGAATTGTCTG 581
101 ArgPheArgThrProGluProAsnHisTyrAsp.....Ph 113
   |||||.....|
582 CGTATCCTTATGGGGAGGCTCTTAATCACCATGACCATGATGATGAATT 631
113 eCysLeuIlePro 117
   |||||.....|
632 TTGCCTTATGCCT 644

```

seq_name: gb_pr:AF187064

seq_documentation_block:

```

LOCUS AF187064 891 bp mRNA 11-JUN-2000
DEFINITION Homo sapiens p75NTR-associated cell death executor (NADE) mRNA,
complete cds.
ACCESSION AF187064
VERSION AF187064.1 GI:8452893
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 891)
AUTHORS Mukai, J., Hachiya, T., Shoji-Hoshino, S., Kimura, M.T., Nadano, D.,
Suvanto, P., Hanaoka, T., Li, Y., Irie, S., Greene, L.A. and Sato, T.A.
TITLE NADE, a p75NTR-associated cell death executor, is involved in
signal transduction mediated by the common neurotrophin receptor
p75NTR
J. Biol. Chem. 275 (23), 17566-17570 (2000)
MEDLINE 20298829
REFERENCE
2 (bases 1 to 891)
AUTHORS Mukai, J., Hachiya, T., Hoshino, S., Kimura, M., Nadano, D.,
Hanaoka, T., Li, Y., Irie, S. and Sato, T.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1999) Otolaryngology/Pathology, Columbia
University, 630 West 168th St., P&S 11-451, New York, NY 10032, USA
FEATURES
source
1..891
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..891
/gene="NADE"
312..647
/gene="NADE"

```



```

115 eutlePro 117
116 117
117 117
118 117
119 117
120 117
121 117
122 117
123 117
124 117
125 117
126 117
127 117
128 117
129 117
130 117
131 117
132 117
133 117
134 117
135 117
136 117
137 117
138 117
139 117
140 117
141 117
142 117
143 117
144 117
145 117
146 117
147 117
148 117
149 117
150 117
151 117
152 117
153 117
154 117
155 117
156 117
157 117
158 117
159 117
160 117
161 117
162 117
163 117
164 117
165 117
166 117
167 117
168 117
169 117
170 117
171 117
172 117
173 117
174 117
175 117
176 117
177 117
178 117
179 117
180 117
181 117
182 117
183 117
184 117
185 117
186 117
187 117
188 117
189 117
190 117
191 117
192 117
193 117
194 117
195 117
196 117
197 117
198 117
199 117
200 117
201 117
202 117
203 117
204 117
205 117
206 117
207 117
208 117
209 117
210 117
211 117
212 117
213 117
214 117
215 117
216 117
217 117
218 117
219 117
220 117
221 117
222 117
223 117
224 117
225 117
226 117
227 117
228 117
229 117
230 117
231 117
232 117
233 117
234 117
235 117
236 117
237 117
238 117
239 117
240 117
241 117
242 117
243 117
244 117
245 117
246 117
247 117
248 117
249 117
250 117
251 117
252 117
253 117
254 117
255 117
256 117
257 117
258 117
259 117
260 117
261 117
262 117
263 117
264 117
265 117
266 117
267 117
268 117
269 117
270 117
271 117
272 117
273 117
274 117
275 117
276 117
277 117
278 117
279 117
280 117
281 117
282 117
283 117
284 117
285 117
286 117
287 117
288 117
289 117
290 117
291 117
292 117
293 117
294 117
295 117
296 117
297 117
298 117
299 117
300 117
301 117
302 117
303 117
304 117
305 117
306 117
307 117
308 117
309 117
310 117
311 117
312 117
313 117
314 117
315 117
316 117
317 117
318 117
319 117
320 117
321 117
322 117
323 117
324 117
325 117
326 117
327 117
328 117
329 117
330 117
331 117
332 117
333 117
334 117
335 117
336 117
337 117
338 117
339 117
340 117
341 117
342 117
343 117
344 117
345 117
346 117
347 117
348 117
349 117
350 117
351 117
352 117
353 117
354 117
355 117
356 117
357 117
358 117
359 117
360 117
361 117
362 117
363 117
364 117
365 117
366 117
367 117
368 117
369 117
370 117
371 117
372 117
373 117
374 117
375 117
376 117
377 117
378 117
379 117
380 117
381 117
382 117
383 117
384 117
385 117
386 117
387 117
388 117
389 117
390 117
391 117
392 117
393 117
394 117
395 117
396 117
397 117
398 117
399 117
400 117
401 117
402 117
403 117
404 117
405 117
406 117
407 117
408 117
409 117
410 117
411 117
412 117
413 117
414 117
415 117
416 117
417 117
418 117
419 117
420 117
421 117
422 117
423 117
424 117
425 117
426 117
427 117
428 117
429 117
430 117
431 117
432 117
433 117
434 117
435 117
436 117
437 117
438 117
439 117
440 117
441 117
442 117
443 117
444 117
445 117
446 117
447 117
448 117
449 117
450 117
451 117
452 117
453 117
454 117
455 117
456 117
457 117
458 117
459 117
460 117
461 117
462 117
463 117
464 117
465 117
466 117
467 117
468 117
469 117
470 117
471 117
472 117
473 117
474 117
475 117
476 117
477 117
478 117
479 117
480 117
481 117
482 117
483 117
484 117
485 117
486 117
487 117
488 117
489 117
490 117
491 117
492 117
493 117
494 117
495 117
496 117
497 117
498 117
499 117
500 117
501 117
502 117
503 117
504 117
505 117
506 117
507 117
508 117
509 117
510 117
511 117
512 117
513 117
514 117
515 117
516 117
517 117
518 117
519 117
520 117
521 117
522 117
523 117
524 117
525 117
526 117
527 117
528 117
529 117
530 117
531 117
532 117
533 117
534 117
535 117
536 117
537 117
538 117
539 117
540 117
541 117
542 117
543 117
544 117
545 117
546 117
547 117
548 117
549 117
550 117
551 117
552 117
553 117
554 117
555 117
556 117
557 117
558 117
559 117
560 117
561 117
562 117
563 117
564 117
565 117
566 117
567 117
568 117
569 117
570 117
571 117
572 117
573 117
574 117
575 117
576 117
577 117
578 117
579 117
580 117
581 117
582 117
583 117
584 117
585 117
586 117
587 117
588 117
589 117
590 117
591 117
592 117
593 117
594 117
595 117
596 117
597 117
598 117
599 117
600 117
601 117
602 117
603 117
604 117
605 117
606 117
607 117
608 117
609 117
610 117
611 117
612 117
613 117
614 117
615 117
616 117
617 117
618 117
619 117
620 117
621 117
622 117
623 117
624 117
625 117
626 117
627 117
628 117
629 117
630 117
631 117
632 117
633 117
634 117
635 117
636 117
637 117
638 117
639 117
640 117
641 117
642 117
643 117
644 117
645 117
646 117
647 117
648 117
649 117
650 117
651 117
652 117
653 117
654 117
655 117
656 117
657 117
658 117
659 117
660 117
661 117
662 117
663 117
664 117
665 117
666 117
667 117
668 117
669 117
670 117
671 117
672 117
673 117
674 117
675 117
676 117
677 117
678 117
679 117
680 117
681 117
682 117
683 117
684 117
685 117
686 117
687 117
688 117
689 117
690 117
691 117
692 117
693 117
694 117
695 117
696 117
697 117
698 117
699 117
700 117
701 117
702 117
703 117
704 117
705 117
706 117
707 117
708 117
709 117
710 117
711 117
712 117
713 117
714 117
715 117
716 117
717 117
718 117
719 117
720 117
721 117
722 117
723 117
724 117
725 117
726 117
727 117
728 117
729 117
730 117
731 117
732 117
733 117
734 117
735 117
736 117
737 117
738 117
739 117
740 117
741 117
742 117
743 117
744 117
745 117
746 117
747 117
748 117
749 117
750 117
751 117
752 117
753 117
754 117
755 117
756 117
757 117
758 117
759 117
760 117
761 117
762 117
763 117
764 117
765 117
766 117
767 117
768 117
769 117
770 117
771 117
772 117
773 117
774 117
775 117
776 117
777 117
778 117
779 117
780 117
781 117
782 117
783 117
784 117
785 117
786 117
787 117
788 117
789 117
790 117
791 117
792 117
793 117
794 117
795 117
796 117
797 117
798 117
799 117
800 117
801 117
802 117
803 117
804 117
805 117
806 117
807 117
808 117
809 117
810 117
811 117
812 117
813 117
814 117
815 117
816 117
817 117
818 117
819 117
820 117
821 117
822 117
823 117
824 117
825 117
826 117
827 117
828 117
829 117
830 117
831 117
832 117
833 117
834 117
835 117
836 117
837 117
838 117
839 117
840 117
841 117
842 117
843 117
844 117
845 117
846 117
847 117
848 117
849 117
850 117
851 117
852 117
853 117
854 117
855 117
856 117
857 117
858 117
859 117
860 117
861 117
862 117
863 117
864 117
865 117
866 117
867 117
868 117
869 117
870 117
871 117
872 117
873 117
874 117
875 117
876 117
877 117
878 117
879 117
880 117
881 117
882 117
883 117
884 117
885 117
886 117
887 117
888 117
889 117
890 117
891 117
892 117
893 117
894 117
895 117
896 117
897 117
898 117
899 117
900 117
901 117
902 117
903 117
904 117
905 117
906 117
907 117
908 117
909 117
910 117
911 117
912 117
913 117
914 117
915 117
916 117
917 117
918 117
919 117
920 117
921 117
922 117
923 117
924 117
925 117
926 117
927 117
928 117
929 117
930 117
931 117
932 117
933 117
934 117
935 117
936 117
937 117
938 117
939 117
940 117
941 117
942 117
943 117
944 117
945 117
946 117
947 117
948 117
949 117
950 117
951 117
952 117
953 117
954 117
955 117
956 117
957 117
958 117
959 117
960 117
961 117
962 117
963 117
964 117
965 117
966 117
967 117
968 117
969 117
970 117
971 117
972 117
973 117
974 117
975 117
976 117
977 117
978 117
979 117
980 117
981 117
982 117
983 117
984 117
985 117
986 117
987 117
988 117
989 117
990 117
991 117
992 117
993 117
994 117
995 117
996 117
997 117
998 117
999 117
1000 117

```

```

seq_name: gb_sts:G72708
seq_documentation_block: 421 bp DNA STS 08-AUG-2001
LOCUS G72708
DEFINITION MARC 4953-4954:991939031:1 SCF - porcine spleen Sus scrofa STS
genomic, sequence tagged site.
ACCESSION G72708
VERSION G72708.1 GI:15146738
KEYWORDS STS
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 421)
AUTHORS Fahrnkruug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L. and
Keele, J.W.
TITLE Single nucleotide polymorphism (SNP) discovery in expressed porcine
genes
JOURNAL Unpublished (2001)
COMMENT
Contact: Freking BA
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel.: 402 762 4278
Fax: 402 762 4173
Email: freking@email.marc.usda.gov
Primer A: GCAATGGGTGAACTCTACT
Primer B: CCAGCAGCAATAGACG
STS size: 500
PCR Profile:
Hotstart: 95 degrees for 15 minutes
Denature: 95 degrees for 30 seconds
Anneal: 56 degrees
Extension: 68 degrees for 2 minutes
Cycles: 32 to 45
Protocol:
Template: 50-200 ng genomic DNA
Primer: each 20 pmol
dNTPs: each 88 uM
Taq Polymerase: 0.25 units (Qiagen HotStar)
Buffer:
Commercially supplied Qiagen HotStar buffer
The STS is derived from PCR amplicons generated from genomic DNA,
sequenced from each end using the amplification primers. The

```



```

LOCUS      HSV351F8      45678 bp      DNA      PRI      23-NOV-1999
DEFINITION Human DNA sequence from cosmid V351F8, between markers DXS366 and
ACCESSION DX587 on chromosome X contains ESTs.
VERSION    270719
KEYWORDS   X.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 45678)
AUTHORS    Whiteley,M.
TITLE      Direct Submission
JOURNAL    Submitted (09-APR-1995) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT     IMPORTANT: This sequence is the entire insert of clone V351F8. The
            true left end of clone V351F8 is at 1 in this sequence. The true
            right end of clone V351F8 is at 45678.
            V351F8 is from the human chromosome X-specific cosmid library.
FEATURES   source
            location/Qualifiers
            1..45678
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="X"
            /map="X"
            /clone="GHC-351F8"
            /clone_lib="SCCV"
            1..70
            /partial
            repeat_region
            /note="Alu repeat: matches 80. .1 of consensus"
            642..930
            /note="Alu repeat: matches 1. .308 of consensus"
            1406..5179
            /note="L1 element fragment"
            5243..5276
            /note="17 copies of 2 mer 82 & conserved"
            5279..5567
            /partial
            repeat_region
            /note="Alu repeat: matches 304. .1 of consensus"
            9975..10369
            /note="MSTA element fragment"
            10399..12017
            /note="MSTAR element fragment"
            11318..11978
            /note="THE1B element fragment"
            11715..11846
            /note="MLTIR element fragment"
            12074..12170
            /note="MSTA element fragment"
            12227..12486
            /note="MSTA element fragment"
            12437..12502
            /note="MSTC element fragment"
            12756..13059
            /partial
            repeat_region
            /note="Alu repeat: matches 308. .1 of consensus"
            14478..14650
            /note="THE1B element fragment"
            14491..14536
            /note="MSTA element fragment"
            14715..14831
            /note="THE1B element fragment"
            15099..15533
            /note="L1 element fragment"
            15534..15626
            /note="31 copies of 3 mer 85 & conserved"
            15566..15619
            /note="3 copies of 18 mer 98 & conserved"
            16352..16436
            /note="L1 element fragment"
            16728..16763
            /note="9 copies of 4 mer 94 & conserved"
            16764..17054
            /partial
            repeat_region
            /note="Alu repeat: matches 308. .1 of consensus"
            19420..19449
            /note="15 copies of 2 mer 87 & conserved"
            complement(19537..20013)
            /note="match: 3' EST N51315 clone 283089"
            20071..20114
            /note="22 copies of 2 mer 98 & conserved"
            20073..20112
            /note="10 copies of 4 mer 100 & conserved"
            20073..20126
            /note="3 copies of 18 mer 87 & conserved"
            23088..23300
            /note="L1 element fragment"
            23311..23490
            /note="MLN1A element fragment"
            23318..23489
            /note="MLN1B element fragment"
            23362..23491
            /note="MLTIC element fragment"
            23507..24407
            /note="L1 element fragment"
            24376..24462
            /note="MSTC element fragment"
            24379..24491
            /note="MSTA element fragment"
            24688..24738
            /note="MSTA element fragment"
            24810..25016
            /note="L1 element fragment"
            25254..25421
            /note="L1 element fragment"
            25439..25696
            /note="MER25 element fragment"
            26284..26343
            /note="MLTID element fragment"
            26428..26491
            /note="MLTID element fragment"
            26465..26518
            /note="MLN1E element fragment"
            27446..27667
            /note="L1 element fragment"
            28954..29247
            /partial
            repeat_region
            /note="Alu repeat: matches 308. .1 of consensus"
            30160..30325
            /note="2 copies of 83 mer 98 & conserved"
            32588..32875
            /note="match: 5' EST H68599 clone 239077"
            complement(32825..33230)
            /note="match: 3' EST H68239 clone 289077"
            35551..35802
            /partial
            repeat_region
            /note="Alu repeat: matches 1. .260 of consensus"
            35804..35839
            /note="18 copies of 2 mer 83 & conserved"
            38625..38705
            /note="MLN1R element fragment"
            38747..38839
            /note="MLTIR element fragment"
            38949..39032
            /note="MLN1E element fragment"
            38965..39076
            /note="MLTID element fragment"
            9135 c 9327 g 12956 t
            BASE COUNT 14260 a
            ORIGIN
            alignment_scores: 138.00 Length: 127
            Quality: 2.000 Gaps: 4
            Ratio: 54.331 Percent Identity: 31.496
            Percent Similarity: 54.331

```

```
alignment_block:
US-09-327-750D-35 x HSV351F8
Align seg 1/1 to: HSV351F8 from: 1 to: 45678

11 LeuthrValGluLysAspLysLysArgGlyGlyLysAlaSerly 27
19408 TATCTTCTAGGAAAGAAAAAAGGAGGAAAAAATCAACAT 19457
27 scInserGluGluProHisHisLeuGluGlu.....ValGluA 41
19458 GGAATGTCCTCCCAAGGAAAAAAGGTTGGGAGAGGCCAGTCGAGA 19507
41 sn.....LysLysProGlyGlyAsn 47
19508 ATGAAGCCCGCGCTTTAGGAGGTGGTGAATACAGGAGCGCTGGAGGAAT 19557
48 ValArgArgLysValArgLeuValProAsnPhaLeuTIPAlaIlePr 64
19558 GTTAAAGGGGTTGGGCTCCACCTGCCCGGGTTTGGAGAGGATGTGCC 19607
64 oAsnArgHisValAspArg.....AsnGluGlyGlyGluAspV 77
19608 CAATAGGCTTCGATAACATGATATAGATGAGATGGAGATGATGATA 19657
77 alGlyArgPhaValValGluGlyThrcluvValLysArgLysThrThrGlu 93
19658 TGAACGGTTCATGAGGATGATGAGAGCTAAGGAGGAAAAATTAGGAA 19707
94 GlnGlnValArgProTyrArgPheArgThrProGluPro..... 107
19708 CTTCAGTTGAGTACAGTCTGCGCATCTTATAGGGGACCCCTCCACCA 19757
108 AsnAsnHisTyrAspPheCysLeuIlePro 117
19758 TGATCATGATGAGTTTGCCTTATGCT 19788

seq_name: gb_sts:G35294

seq_documentation_block: 477 bp DNA STS 02-OCT-1997
LOCUS G35294
DEFINITION human STS SHGC-37409, sequence tagged site.
ACCESSION G35294
VERSION G35294.1 GI:2459462
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 477)
Myers, R.M.
Unpublished (1997)
Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 415/725/9687
Fax: 415/725/9689
Email: myers@shgc.stanford.edu
Primer A: AACATCTTTCCATGAAGATGATG
Primer B: CTTTGGCATTTCTTCCAA
STS size: 106
PCR Profile:
Initial incubation: 95 degrees C for 10 minutes
Denaturation: 94 degrees C for 15 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 23 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9600
Protocol:
```

```
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
AmpliTaq Gold Polymerase: 0.07 units/ul
Total Vol: 5 ul

Buffer: MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3

Prepared with primer pairs provided by Sandoz, derived from N51315
-- Washington University/Merck EST sequence.
FEATURES
source
1..477
/organism="Homo sapiens"
/db_xref="taxon:9606"
44..149
primer_bind'
44..67
primer_bind'
complement(130..149)
BASE COUNT 153 a 115 c 79 g 130 t
ORIGIN

alignment_scores:
Quality: 120.50 Length: 83
Ratio: 2.410 Gaps: 2
Percent Similarity: 60.241 Percent Identity: 36.145

alignment_block:
US-09-327-750D-35 x G35294/rev ..
Align seg 1/1 to reverse of: G35294 from: 1 to: 477

42 LysLysProGlyGlyAsnValArgArgLysValArgLeuValProAs 58
474 CAGGAGCGCTGGAGAAATGTTAAAGGGGTTTGGGCTCCACCTGCCCGGG 425
58 nPheLeuTIPAlaIleProAsnArgHisValAspArg.....A 71
424 TTTTGGAGAGGAGTGGCCCAATAGGCTTTCGATAACATGATGATAG 375
71 snGluGlyGlyGluAspValGlyArgPheValValGlnGlyThrGluVal 87
374 ATGGACATGAGATGATATGGACCGTTTCATGGAGGAGATGAGAGCTA 325
88 LysArgLysThrThrGluGlnValArgProTyrArgArgPheArgTh 104
324 AGGAGAAATAGGGAACCTTCAGTTGAGGTACAGTCTGCGCATTTCTAT 275
104 rProGluPro.....AspAsnHisTyrAspPheCysLeuIlePro 117
274 AGGGAGCGCTCCCTCACCATGATCATCATGATGAGTTTGCCTTATGCT 226

seq_name: gb_pr:HS714B7

seq_documentation_block:
LOCUS HS714B7 98274 bp DNA PRI 12-DEC-1999
DEFINITION Human DNA sequence from clone CTA-714B7 on chromosome 22q12.2-13.2
Contains pseudogene similar to part of COX7B (cytochrome c oxidase
subunit VIIb), a novel pseudogene, ESTs, STS and GSSs, complete
sequence.
ACCESSION Z99755
VERSION Z99755.1 GI:3036782
KEYWORDS HTG; COX7B; cytochrome c oxidase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 98274)
Connor, R.
Direct Submission
```


508 AGGCGTTTCCGAACCCCGGAACCTGACATCATTTAGCTTTTGCCTCAT 557

117 ePro 118
|||||
558 ACCT 561

seq_name: gb_est2.BG083261

seq_documentation_block: 792 bp. mRNA EST 26-JAN-2001
LOCUS BG083261
DEFINITION H3086C08-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
H3086C08 5', mRNA sequence.
ACCESSION BG083261
VERSION BG083261.1 GI:12565829
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 792)
AUTHORS Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A., Tanaka
T.S., Carter, M.G. and KO, M.S.H.
TITLE Verification and initial annotation of NIA mouse 15K cDNA clone set
JOURNAL Unpublished (2001)
COMMENT Other_ESTs: H3086C08-3
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdnaelgscn.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit <http://lgsun.grc.nia.nih.gov/cDNA/15k.html> for details.
Plate: H3086 Row: C Column: 08
Seq primer: -21M13 Reverse
High quality sequence stop: 792
POLYA-No.

FEATURES
source Location/Qualifiers
1..792
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="H3086C08"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_stage="Clones arrayed from a variety of cDNA libraries"
/lab_host="DH10B"
/note="vector: pSPORT1; Site_1: SalI; Site_2: NotI; This clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."
BASE COUNT 247 a 153 c 210 g 179 t 3 others
ORIGIN

alignment_scores:
Quality: 546.00 Length: 118
Ratio: 4.964 Gaps: 0

Percent Similarity: 93.220 Percent Identity: 87.288
alignment_block:
US-09-327-750D-34 x BG083261 ..
Align seg 1/1 to: BG083261 from: 1 to: 792
1 MetAlaSerLysValLysGlnValLleLeuAspLeuThrValGluLysAs 17
|||||
190 ATGCATCCAAATTTAAACAAGTCATATCTGGATCTCACTGTGGAGAAGA 239
|||||
17 pLysLysAsnLysGlyLysAlaSerLysGlnSerGluGluGlu 34
|||||
240 CAAAAAGACAAAAAGGTGGGAAGGCTCCAAACAAAGTCAAGAAAGAAC 289
|||||
34 erHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 50
|||||
290 CCCACCATCTCGAAGAGGTTCAAAACAAGAGCCCTGGGGGAAATGTCGGA 339
|||||
51 ArgLysValArgArgLeuValProAsnPhelLeuTrpAlaLleProAsnAr 67
|||||
340 AGGAAAGTCAGCGGACTTTGTGCTTAACCTTCTCTGGCCATACCAAAATAG 389
|||||
67 gHisValAspHisSerGluGlyGlyGluValGlyArgPheValGlyG 84
|||||
390 GCATGTTGATCGCAATGAAGGGGAGAGGATGTTGGGAGATTGTAGTGC 439
|||||
84 InValMetGluAlaLysArgHisSerLysGluGlnGlnMetArgProTyr 100
|||||
440 AGGGAACAGAGTCAAGAGAAGACTACGGAGCAGCAGCTGAGGCTTAC 489
|||||
101 ThrArgPheArgThrProGluProAspAsnHisTyrAspPheCysLeuIl 117
|||||
490 AGGCGTTTCCGAACCCCGGAACCTGACAATCATTCAGACTTTTGCCTCAT 539
|||||
117 ePro 118
|||||
540 ACCT 543

seq_name: gb_est1:AA473525

seq_documentation_block: 458 bp mRNA EST
LOCUS AA473525
DEFINITION v978a09.r1 Barstead MFLRB1 Mus musculus. cDNA clone IMAGE:872056 5', similar to SW:HG74_HUMAN Q00994 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74. ;, mRNA sequence.

ACCESSION AA473525
VERSION AA473525.1 GI:2201752
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 458)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soates, B., Wilson, R. and Waterston, R.

TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Maria M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:511536
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 417.

DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 image.llnl.gov/image/html/resources.shtml
 Seq primer: -40UP from Gibco
 High quality sequence stop: 422.

FEATURES

source
 1. .612
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2849369"
 /clone_lib="NCI-CGAP_Lym12"
 /tissue_type="lymphoma, follicular mixed small and large
 cell"
 /lab_host="DH10B"
 /note="Organ: lymph node; Vector: pCMV-SPORT6; Site_1:
 SalI; Site_2: NotI; Cloned unidirectionally. Primer:
 oligo dh. Average insert size 1.25 kb. Life Technologies
 catalog #: 11547-015"
 BASE COUNT 169 a 145 c 120 g 178 t
 ORIGIN

alignment_scores:
 Quality: 609.00 Length: 111
 Ratio: 5.486 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-327-750D-13 x AW512400/rev ..

Align seg 1/1 to reverse of: AW512400 from: 1 to: 612

1 MetAlaAsnIleHisGlnGluAsnGluGluMetGluGlnProMetGlnAs 17
 |||||
 581 ATGGCAAAATATTCACCGAGAAACGAGAGATGGAGCGCTATGCAGAA 532
 17 nGlyGluGluAspArgProLeuGlyGlyGlyGlyHisGlnProAla 34
 |||||
 531 TGGAGAGGAGACCGCCCTTTGGAGAGGTTGAAGGCCACCGCCCTGCAG 482
 34 lyAsnArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAla 50
 |||||
 481 GAAATCGACGGGACAGCGCTCGCGACTTCCCTTAATTTTCATGGGCC 432
 51 IleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspMe 67
 |||||
 431 ATACCAATAGGCAGATCAATGATGGATGGGTGGAGATGGAGATGATAT 382
 67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84
 |||||
 381 GGAATATTTCATGGAGAGATGAGAGAAATCAGAGAAACTTAGGGAGC 332
 84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHis 100
 |||||
 331 TGCAGTTGAGCAATTGTCGCTATCCCTATGGGGAGCTCTCTAATCAC 282
 101 HisAspHisHisAspGluPheCysLeuMetPro 111
 |||||
 281 CATGACCATCATGATCAATTTTGCCTTATGCCT 249

seq_name: gb_est1:AI193112

seq_documentation_block:

LOCUS AI193112 615 bp mRNA EST 29-OCT-1998
 DEFINITION qe69g09_x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
 IMAGE:1744288 3' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.

ACCESSION AI193112
 VERSION AI193112.1 GI:3744321
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 615)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (image.llnl.gov) for further information.
 INSERT Length: 845 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 445.
 Location/Qualifiers
 1. .615

FEATURES

source
 1. .615
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1744288"
 /clone_lib="Soares_fetal_lung_NbHL19W"
 /dev_stage="19 weeks"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
 modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer
 [5'-TGTTACCAATCTGAAGTGGAGCGCCCAATTTTCTTTT-3'],
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT7T3 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M.Patima Bonaldo. This library was constructed
 from the same fetus as the fetal heart library, Soares
 fetal heart NbHL19W."
 BASE COUNT 169 a 145 c 121 g 180 t
 ORIGIN

alignment_scores:
 Quality: 609.00 Length: 111
 Ratio: 5.486 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-327-750D-13 x AI193112/rev ..

Align seg 1/1 to reverse of: AI193112 from: 1 to: 615

1 MetAlaAsnIleHisGlnGluAsnGluGluMetGluGlnProMetGlnAs 17
 |||||
 584 ATGGCAAAATATTCACCGAGAAACGAGAGATGGAGCGCTATGCAGAA 535
 17 nGlyGluGluAspArgProLeuGlyGlyGlyGlyHisGlnProAla 34
 |||||
 534 TGGAGAGGAGACCGCCCTTTGGAGAGGTTGAAGGCCACCGCCCTGCAG 485
 34 lyAsnArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAla 50
 |||||
 484 GAAATCGACGGGACAGCGCTCGCGACTTCCCTTAATTTTCATGGGCC 435
 51 IleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspMe 67
 |||||
 434 ATACCAATAGGCAGATCAATGATGGATGGGTGGAGATGGAGATGATAT 385
 67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84
 |||||
 384 GGAATATTTCATGGAGAGATGAGAGAAATCAGAGAAACTTAGGGAGC 335
 84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHis 100
 |||||
 334 TGCAGTTGAGCAATTGTCGCTATCCCTATGGGGAGCTCTCTAATCAC 285
 101 HisAspHisHisAspGluPheCysLeuMetPro 111

```

|||||
284 CATGACCATCATGATGAATTTTGCCTTATGCCT 252
seq_name: gb_est1:AA576958

seq_documentation_block:
LOCUS AA576958 620 bp mRNA 12-SEP-1997
DEFINITION nm82d04.s1 NCI-CGAP-Co9 Homo sapiens cDNA clone IMAGE:1074727 3'
similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74
(HUMAN); mRNA sequence.
ACCESSION AA576958
VERSION AA576958.1 GI:2354432
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 620)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs@remail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 725 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 339.
Location/Qualifiers
1..620
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1074727"
/clone_lib="NCI-CGAP-Co9"
/tissue_type="colon tumor RER+"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
RER+ colon tumor, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is not normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo (Soares4
)."
```

```

BASE COUNT 169 a 144 c 116 g 191 t
ORIGIN
```

```

alignment_scores:
Quality: 609.00 Length: 111
Ratio: 5.486 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

```

alignment_block:
US-09-327-750D-13 x AA576958/rev ..
```

```

Align seg 1/1 to reverse of: AA576958 from: 1 to: 620
```

```

1 MetalaasnleHisGlnGluAsnGluGluMetGluGlnProMetGlnAs 17
|||||
```

```

602 ATGCCAATAATTCACGAGAAACGAAGAGATGGAGCGCTATGSCAGAA 553
```

```

17 nGlyGluGluAspArgProLeuGlyGlyGlyGluGlyHisGlnProAlaG 34
|||||
```

```

552 TGGAGAGAGAGACGCCCTTTGGGAGGAGGGAAGGCCACCAGCCTCGAG 503
```

```

34 lyAsnArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAla 50
|||||
502 GAAATCGACGGGACAGCGCTCGCCGACTTCCCTTAATTTTCGATGGGCC 453

51 IleProAsnArgGlnIleAsnAspGlyMetGlyVclYAspGlyAspAspMe 67
|||||
452 ATACCCCAATAGGAGATCATGATGGATGGTGGAGATGGAGATGATAT 403

67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84
|||||
402 GGAATATTCATGGAGAGATGAGAGAAATCAGAGAAAACCTTAGGGAGC 353

84 euGlnIleuArgAsnCysLeuArgIleuMetGlyGluLeuSerAsnHis 100
|||||
352 TGCAGTTGAGGAATGTCTCGTATCCTTATGGGGAGAGCTCTCTAATCAC 303

101 HisAspHisHisAspGluPheCysLeuMetPro 111
|||||
302 CATGACCATCATGATGAATTTTGCCTTATGCCT 270
```

```
seq_name: gb_est1:AI929106
```

```
seq_documentation_block:
```

```

LOCUS AI929106 662 bp mRNA 23-AUG-1999
DEFINITION au65b10.y1 Schneider fetal brain 00004 Homo sapiens cDNA clone
IMAGE:2519611 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
KD PROTEIN HGR74 (HUMAN); mRNA sequence.
ACCESSION AI929106
VERSION AI929106.1 GI:5665070
KEYWORDS EST.
SOURCE human.
```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 662)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
```

```

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40RP from Gibco
High quality sequence stop: 437.
Location/Qualifiers
1..662
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2519611"
/clone_lib="Schneider fetal brain 00004"
/sex="male"
/tissue_type="frontal lobe"
/dev_stage="5 months post-conception"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript SK (Stratagene);
Site 1: SstI; Site 2: XhoI; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
5'-GAGAGAGAGAGAGCTCAAGGATCTCTAATTAATTAATCCCCCCCCCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGAGCTCGAGTTTCTTTT-3'. The library was
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
```

```

FEATURES
SOURCE
```

102(b)

11 (mouse)

polyA_signal 790..795
/note="putative"
polyA_site 809
/note="putative"
BASE COUNT 208 a 211 c 227 g 163 t
ORIGIN

alignment_scores:
Quality: 688.00 Length: 124
Ratio: 5.548 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.387
alignment_block:
US-09-327-750D-12 x AK004531

Align seg 1/1 to: AK004531 from: 1 to: 809
1 MetAlaAsnValHisGlnGluAsnGluMetGluGlnProLeuGlnAs 17
181 ATGGCCATATGTCACACGAGAAACGAGAGATGGAGCAGCCCTGCAGAA 230
17 nGlyGluGluAspArgProValGlyGlyGlyGlyGlnHisGlnProAlaG 34
231 TGGACAGGAAGACCGCCCTGTGGAGGAGGTGAGGCGCCACCGCTGCTG 280
34 LyAsnAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisAsnHisArg 50
281 CAACACAAACAAACAAACAAACCAACCAACCAACCAACCAACCAACCA 330
51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTTPAlaLalleProAsnAr 67
331 GCCCAGGCTCGCGGACTTGCCTTAACCTCCGATGGGCGCATTCACCAACAG 380
67 gGlnMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheM 84
381 GCAGATGAATGACGGTGGGTGGAGATGGAGATGATGGAATGTCA 430
84 etGluGluMetArgGluLeuArgArgLysLeuArgGluLeuLeuArg 100
431 TGGAGGAGATGAGAGAGATCCGAGAAAGCTTAGGGAGCTACAGCTGAGA 480
101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisAsnHis 117
481 AATTGCTACGGATCTTATGGGGAGCTGTCTAACCAACCAACCAACCA 530
117 sAspGluPheCysLeuMetPro 124
531 TGATGAATTCGCTTATGCGCT 552
seq_name: gb_est1:AA253897

seq_documentation_block:
LOCUS AA253897 468 bp mRNA EST
DEFINITION MY31e10.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA
clone IMAGE:697482 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
13.0 KD PROTEIN HG874 (HUMAN);, mRNA sequence.
ACCESSION AA253897
VERSION AA253897.1 GI:1888253
KEYWORDS EST.
SOURCE mouse house.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 468)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:431042
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 455.
Location/Qualifiers
1..468
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:697482"
/sex="mixed"
/tissue_type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTCAATCTCAAGTGGAGCGCCCTTTTTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[GTTCGATTCGGTAC], digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library constructed by Bob Barstead."

FEATURES

SOURCE

BASE COUNT 140 a 126 c 120 g 82 t
ORIGIN

alignment_scores:

Quality: 687.00 Length: 124
Ratio: 5.540 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 97.581
alignment_block:
US-09-327-750D-12 x AA253897

Align seg 1/1 to: AA253897 from: 1 to: 468

1 MetAlaAsnValHisGlnGluAsnGluMetGluGlnProLeuGlnAs 17
49 ATGGCCATATGTCACACGAGAAACGAGAGATGGAGCAGCCCTGCAGAA 98
17 nGlyGluGluAspArgProValGlyGlyGlyGlyGlnHisGlnProAlaG 34
99 TGGACAAACAAACAAACCAACCAACCAACCAACCAACCAACCAACCA 148
34 LyAsnAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisAsnHisArg 50
149 CAACACAAACAAACCAACCAACCAACCAACCAACCAACCAACCAACCA 198
51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTTPAlaLalleProAsnAr 67
199 GCCCAGGCTCGCGGACTTGCCTTAACCTCCGATGGGCGCATTCACCAACAG 248
67 gGlnMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheM 84
249 GCAGATGAATGACGGTGGGTGGAGATGGAGATGATGGAATGTTC 298
84 etGluGluMetArgGluLeuArgArgLysLeuArgGluLeuLeuArg 100
299 TGGAGGAGATGAGAGATCCGAGAAAGCTTAGGGAGCTACAGCTGAGA 348
101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisAsnHisAsnHis 117
349 AATTGCTACCGCTTATGGGGAGCTGTCTAACCAACCAACCAACCA 398
117 sAspGluPheCysLeuMetPro 124

11(mouse)

US-09-327-750D-12 x BG088461

Align seg 1/1 to: BG088461 from: 1 to: 599

```
1 MetAlaAsnValHisGlnGluAsnGluMetGluGlnProLeuGlnAs 17
175 ATGCCCAATGTCCACCAAGAAAGAGAGATGAGCAGCCCTGCAGAA 224
17 nGlyGluGluAspArgProValGlyGlyGlyGluGluGlnProAlaG 34
225 TGGACAGAGAGACCGCCTGTGGGAGGAGGTGAGGGCCACAGCCTGCTG 274
34 lyAsnAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisAsnHis 50
275 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 324
51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnAr 67
325 GGCCAGGCTGCCGACTTGGCCCTTAACCTCCGATGGGCCATTGCCAACAG 374
67 gGlnMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheM 84
375 GCAGATGAATGACCGGTGGGTGGAGATGGAGATGATGGAATGTTCA 424
84 etGluGluMetArgGluIleArgArgLysLeuArgGluGluGlnLeuArg 100
425 TGGAGGAGATGAGAGATCCGGAGAAAGCTTAGGAGCTACAGCTGAGA 474
101 AsnCysLeuArgLleLeuMetGlyGluLeuSerAsnHisAsnHis 117
475 AATTGCTACGCATCTTATGGGGAGCTGTCTAACCAACACGATCACA 524
117 sAspGluPheCysLeuMetPro 124
525 TGATGAATTCCTGCTTATGCCT 546
```

seq_name: gb_est1:AA272375

seq_documentation_block: 612 bp mRNA EST 26-MAR-1997
LOCUS AA272375
DEFINITION vb62g07.r1 Barstead mouse pooled organs MPLRB4 Mus musculus CDNA
clone IMAGE:761628 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.

ACCESSION AA272375
VERSION AA272375.1 GI:1910706
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus-musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 612)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE The WashU-HMMI Mouse EST Project
JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project
WashU-HMMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:462548

Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 507.

FEATURES Location/Qualifiers
source 1..612
/organism="Mus musculus"

The

/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:761628"
/clone_lib="Barstead mouse pooled organs MPLRB4"
/sex="mixed"
/tissue_type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5'
TGTTACGAATCTGAAGTGGAGCGCCGCTTTTCTTTTCTTTTCTTTTCTTTT
3'); double-stranded cDNA was ligated to Eco RI adaptors
[CTTGGATTGGTACCT], digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Barstead."

BASE COUNT 168 a 167 c 175 g 102 t
ORIGIN

alignment_scores:
Quality: 688.00 Length: 124
Ratio: 5.548 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.387

alignment_block:

US-09-327-750D-12 x AA272375

Align seg 1/1 to: AA272375 from: 1 to: 612

```
1 MetAlaAsnValHisGlnGluAsnGluMetGluGlnProLeuGlnAs 17
175 ATGCCCAATGTCCACCAAGAAAGAGATGAGCAGCCCTGCAGAA 225
17 nGlyGluGluAspArgProValGlyGlyGlyGluGluGlnProAlaG 34
225 TGGACAGGAGACCGCCTGTGGGAGGAGGTGAGGGCCACAGCCTGCTG 275
```

```
34 lyAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisAsnHisArgArg 50
275 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 325
```

```
51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnAr 67
325 GGCCAGGCTGCCGACTTGGCCCTTAACCTCCGATGGGCCATTGCCAACAG 375
```

```
67 gGlnMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheM 84
375 GCAGATGAATGACCGGTGGGTGGAGATGGAGATGATGGAATGTTCA 425
```

```
84 etGluGluMetArgGluIleArgArgLysLeuArgGluGluGlnLeuArg 100
425 TGGAGGAGATGAGAGATCCGGAGAAAGCTTAGGAGCTACAGCTGAGA 475
```

```
101 AsnCysLeuArgLleLeuMetGlyGluLeuSerAsnHisAsnHis 117
475 AATTGCTACGCATCTTATGGGGAGCTGTCTAACCAACACGATCACA 525
```

```
117 sAspGluPheCysLeuMetPro 124
```

```
525 TGATGAATTCCTGCTTATGCCT 547
```

seq_name: gb_est2:W81757

seq_documentation_block:

LOCUS W81757 616 bp mRNA EST 12-SEP-1996
DEFINITION me95d06.r1 Soares mouse embryo NDME13.5 14.5 Mus musculus CDNA
clone IMAGE:403307 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.

ACCESSION W81757
VERSION W81757.1 GI:1392776
KEYWORDS EST.
SOURCE house mouse.

11(mouse)

```
US-09-327-750D-12 x W46041
Align seg 1/1 to: W46041 from: 1 to: 492

1 MetAlaAsnValHisGlnGluAsnGluMetGluGlnProLeuGlnAs 17
13 ATGGCCAAATGTCACCAAGAAACGAAGAGATGGAGCAGCCCTGCAGAA 62
17 nGlyGluGluAspArgProValGlyGlyGlyGlyGlyGlnProAlaG 34
63 TGGACAGGAAGACCCGCTGTGGGAGGAGTGGAGGCCACCAAGCTGCTG 112
34 lYAsnAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisArgArg 50
113 CAACACACACACACACACACACACACACACACACACACACACAC 162
51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnAr 67
163 GGCCAGGCTGCGCGAGCTTCCGCTTCCGATGGGCGCATTCACCAACAG 212
67 glnMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheM 84
213 GCAGATGAATGACGGGTTGGGTGGAGATGGAGATGATGGAATGTTCA 262
84 etGluGluMetArgGluIleArgArgLysLeuArgGluLeuGlnLeuArg 100
263 TGGAGGAGATGAGAGATGCGGAGAAAGCTTAGGGAGCTACAGCTGAGA 312
101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisAsnHisAspHisH 117
313 AATTGTCTACGATCCTTATGGGAGAGCTGTAAACACACACAGATCACCA 362
117 saspGluPheCysLeuMetPro 124
363 TGATGAATTCGCTTATGCCT 384

seq_name: gb_est1:AI006575
```

```
seq_documentation_block:
LOCUS AI006575 538 bp mRNA 12-JUN-1998
DEFINITION ue15e06.y1 Sugano mouse embryo mewa Mus musculus cDNA clone
IMAGE:1480450 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
KD PROTEIN HGR74 (HUMAN); mRNA sequence.
ACCESSION AI006575
VERSION AI006575.1 GI:3216184
SOURCE EST.
ORGANISM house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 538)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:928806
Seq primer: primer name ambiguous
High quality sequence stop: 433.
Location/Qualifiers
1..538
/organism="Mus musculus"
```

```
seq_documentation_block:
LOCUS BE334866 542 bp mRNA 14-JUL-2000
DEFINITION us90all.y1 Soares mammary gland_NMLMG Mus musculus cDNA clone
IMAGE:3325628 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
KD PROTEIN HGR74 (HUMAN); mRNA sequence.
ACCESSION BE334866
VERSION BE334866.1
SOURCE EST.
ORGANISM house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 538)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:928806
Seq primer: primer name ambiguous
High quality sequence stop: 433.
Location/Qualifiers
1..538
/organism="Mus musculus"
```

```
seq_name: gb_est1:BE334866
seq_documentation_block:
LOCUS BE334866 542 bp mRNA 14-JUL-2000
DEFINITION us90all.y1 Soares mammary gland_NMLMG Mus musculus cDNA clone
IMAGE:3325628 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
KD PROTEIN HGR74 (HUMAN); mRNA sequence.
ACCESSION BE334866
```

```
seq_name: gb_est1:BE334866
seq_documentation_block:
LOCUS BE334866 542 bp mRNA 14-JUL-2000
DEFINITION us90all.y1 Soares mammary gland_NMLMG Mus musculus cDNA clone
IMAGE:3325628 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
KD PROTEIN HGR74 (HUMAN); mRNA sequence.
ACCESSION BE334866
```

101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisAspHisH117
 |||||
 478 AATTGCTACGCATCTTATGGGGAGCTGTCTAACCAACAGATCACCA 527
 |||||
 117 sAspGluPheCysLeuMetPro 124
 |||||
 528 TGATGAATTCGCTTATGCCT 549
 |||||

seq_name: gb_est1.A1152323

seq_documentation_block:
 LOCUS A1152323 430 bp mRNA EST
 DEFINITION uq79a02.r1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
 IMAGE:1477034 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.

ACCESSION A1152323

VERSION A1152323.1 GI:3680792

KEYWORDS EST

SOURCE house mouse

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 430)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Marra M., Hallier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:925390

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 419.

FEATURES

source

1. .430

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:1477034"

/clone_lib="Soares_mammary_gland_NMLMG"

/sex="female (lactating)"

/tissue_type="mammary gland"

/lab_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; 1st strand cDNA was prepared from mammary
 gland tissue from a lactating female, and was then primed
 with a Not I - oligo(dT) primer. Double-stranded cDNA was
 ligated to Eco RI adaptors (Pharmacia), digested with Not
 I and cloned into the Not I and Eco RI sites of the
 modified pT7T3 vector. Library is normalized. Library
 was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

ORIGIN

135 a 111 c 113 g 71 t

alignment_scores:

Quality: 681.00

Ratio: 5.537

Percent Similarity: 99.194

Percent Identity: 97.581

alignment_block:

US-09-327-750D-12 x A1152323

Align seg 1/1 to: A1152323 from: 1 to: 430

FEATURES

source

1. .503

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:1515440"

/clone_lib="Soares_mammary_gland_NMLMG"

/sex="female (lactating)"

/tissue_type="mammary gland"

/lab_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; 1st strand cDNA was prepared from mammary
 gland tissue from a lactating female, and was then primed
 with a Not I - oligo(dT) primer. Double-stranded cDNA was
 ligated to Eco RI adaptors (Pharmacia), digested with Not

1 MetAlaAsnValHisGlnGluAsnGluGluMetGluGlnProLeuGlnAs 17
 |||||
 54 ATGGCAATGTCACAGAAACCAAGAGATGGAGCAGCCCTGCAGAA 103
 |||||
 17 nGlyGluGluAspArgProValGlyGlyGlyGlyGluGlnProAlaG 34
 |||||
 104 TGGACAGGAAGACCGCCCTGTGGATGGAGGTGAGGCCACACAGCTGTG 153
 |||||
 34 lYAsnAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisHisArgArg 50
 ::::|
 154 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 203
 |||||
 51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTTPAlaIleProAsnAr 67
 |||||
 204 GGCCAGGCTCGCCGACTTCCCTTAACCTCCGATGGGCCATTCCCAACAG 253
 |||||
 67 gGlnMetAsnAspGlyLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 84
 |||||
 254 GCAGATGAATGACGGTGGTGGAGATGGAGATGATATGGAATGTTCA 303
 |||||
 84 etGluGluMetArgGluIleArgArgLysLeuArgGluGlnLeuArg 100
 |||||
 304 TGGAGGAGATGAGAGATCGGAGAAAGCTTAGGGAGCTACAGCTGAGA 353
 |||||
 101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisAspHisH 117
 |||||
 354 AATTGCTACGCATCTTATGGGGAGCTGTCTAACCAACAGATCACCA 403
 |||||

seq_name: gb_est1.AW908751

seq_documentation_block:

LOCUS AW908751 503 bp mRNA EST 25-MAY-2000

DEFINITION uf57a05.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone

IMAGE:1515440 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0

KD PROTEIN HGR74 (HUMAN);, mRNA sequence.

ACCESSION AW908751

VERSION AW908751.1 GI:8073984

KEYWORDS EST

SOURCE house mouse

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 503)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Other_ESTS: uf57a05.x1

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:942292

Seq primer: -40RP from Gibco

High quality sequence stop: 468.

Location/Qualifiers

1. .503

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:1515440"

/clone_lib="Soares_mammary_gland_NMLMG"

/sex="female (lactating)"

/tissue_type="mammary gland"

/lab_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; 1st strand cDNA was prepared from mammary
 gland tissue from a lactating female, and was then primed
 with a Not I - oligo(dT) primer. Double-stranded cDNA was
 ligated to Eco RI adaptors (Pharmacia), digested with Not

102(6)

US-09-327-750d-12 x BG088461

Align seg 1/1 to: BG088461 from: 1 to: 599

1 MetAlaAsnValHisGlnGluAsnGluMetGluGlnProLeuGlnAs 17
 175 ATGGCCCAATGTCCACAGAAACGAAGAGATGGAGCCCTGCAGAA 224
 17 nGlyGluGluAspArgProValGlyGlyGlyGlyGlyGlyHisGlnProAlaG 34
 225 TGGACAGGAAGACCCCTGTGGGAGAGGTGAGGGCCACCAAGCTGCTG 274
 34 lYAsnAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisAsnHisArg 50
 275 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAGA 324
 51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTrrPAlaIleProAsnAr 67
 325 GGCCAGGCTCGCCGACTTGCCTTAACCTCCGATGGGCCATTCCCAACAG 374
 67 gGlnMetAsnAspGlyLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 84
 375 GCAGATGAATGACGGGTGGGTGGAGATGGAGATGATGGAATGTTCA 424
 84 etGluGluMetArgGluLeuArgArgLysLeuArgGluLeuGlnLeuArg 100
 425 TGGAGGAGATGAGAGAGATCGGAGAAAGCTTACGGAGCTACAGCTGAGA 474
 101 AsnCysLeuArgGluLeuMetGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 117
 475 AATTGTCTACGCATCTTATGGGGAGCTGTCTAACCAACCAACGATCACCA 524
 117 sAspGluPheCysLeuMetPro 124
 525 TGATGAATTCGCCTTATGCCT 546

seq_name: gb_est1:AA272375

seq_documentation_block:

LOCUS AA272375 612 bp mRNA EST 26-MAR-1997
 DEFINITION vb69707.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA
 clone IMAGE:761628 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
 13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.

ACCESSION AA272375
 VERSION AA272375.1 GI:1910706
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 612)
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
 TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:462548

Seq primer: -28ml3 rev2 Et from Amersham

High quality sequence stop: 507.

Location/Qualifiers

1. 612

/organism="Mus musculus"

/strain="FVB/N"
 /db_xref="taxon:10090"
 /clone_lib="IMAGE:761628"
 /clone_lib="Barstead mouse pooled organs MPLRB4"
 /sex="mixed"
 /tissue_type="pooled organs"
 /dev_stage="7 day"
 /lab_host="DH10B"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTACGAATCTGAAGTGGAGCGGCCCTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [GTGTGATTCGGTACC], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed by Bob Barstead."

BASE COUNT 168 a 167 c 175 g 102 t
 ORIGIN

alignment_scores:

Quality: 688.00 Length: 124
 Ratio: 5.548 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 98.387

alignment_block:

US-09-327-750D-12 x AA272375

Align seg 1/1 to: AA272375 from: 1 to: 612

1 MetAlaAsnValHisGlnGluAsnGluMetGluGlnProLeuGlnAs 17
 175 ATGGCCCAATGTCCACAGAAACGAAGAGATGGAGCCCTGCAGAA 225
 17 nGlyGluGluAspArgProValGlyGlyGlyGlyGlyGlyHisGlnProAlaG 34
 226 TGGACAGGAAGACCCCTGTGGGAGAGGTGAGGGCCACCAAGCTGCTG 275
 34 lYAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisAsnHisArg 50
 276 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAGA 325
 51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTrrPAlaIleProAsnAr 67
 326 GGCCAGGCTCGCCGACTTGCCTTAACCTCCGATGGGCCATTCCCAACAG 375
 67 gGlnMetAsnAspGlyLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 84
 376 GCAGATGAATGACGGGTGGGTGGAGATGGAGATGATGGAATGTTCA 425
 84 etGluGluMetArgGluLeuArgArgLysLeuArgGluLeuGlnLeuArg 100
 426 TGGAGGAGATGAGAGAGATCGGAGAAAGCTTACGGAGCTACAGCTGAGA 475
 101 AsnCysLeuArgGluLeuMetGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 117
 476 AATTGTCTACGCATCTTATGGGGAGCTGTCTAACCAACCAACGATCACCA 525
 117 sAspGluPheCysLeuMetPro 124
 526 TGATGAATTCGCCTTATGCCT 547

seq_name: gb_est2:W81757

seq_documentation_block:

LOCUS W81757 616 bp mRNA EST 12-SEP-1996
 DEFINITION me95d06.r1 Soares mouse embryo NM0E13.5 14.5 Mus musculus cDNA
 clone IMAGE:403307 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
 13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.

ACCESSION W81757
 VERSION W81757.1 GI:1392776
 KEYWORDS EST.
 SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 616)

REFERENCE
AUTHORS Marra M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T., Geisel S., Kucaba T., Lacy M., Le M., Martin J., Morris M., Schellenberg K., Steptoe M., Tan F., Underwood K., Moore B., Theising B., Wylie T., Lennon G., Soares B., Wilson R. and Waterston R.

TITLE
The WashU-HMI Mouse EST Project

JOURNAL
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:247075

FEATURES
source Putative full length read
vector: to vector length is 617
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 492.
Location/Qualifiers
1..616
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:403307"
/clone_lib="Soares mouse embryo NDBE13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="vector: pT7T3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGGAGCGCGCGAAATTTTTTTTTTTTTTTT 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 159 a 159 c 159 g 139 t

ORIGIN

alignment_scores:
Quality: 688.00 Length: 124
Ratio: 5.548 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.387

alignment_block:
US-09-327-750d-12 x W81757
Align seg 1/1 to: W81757 from: 1 to: 616

1 MetAlaAsnValHisGlnGluAsnGluMetGluGlnProLeuGlnAs 17
7 ATGGCCAATGTCCACAGGAAAACGAAGATGGAGCAGCCCTGCAGAA 56
17 nGlyGluGluAspArgProValGlyGlyGlyGlyHisGlnProAlaG 34
57 TGGACAGGAAGACCGCCCTGTGGAGAGGTGAGGCCACCGCTGCTG 106
34 lYAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisHisArgArg 50
107 CAAC 156

51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTTPAlaIleProAsnAr 67
157 GCCCAGGCTCGCCGACTTGCCTTAACTTCCGATGGCCATTTCCCAACAG 206
67 gGlnMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheM 84
207 GCAGATGAATGACGGTGGGTGGAGATGGAGATGATATGGAATGTTCA 256
84 etGluGluMetArgGluIleArgArgLysLeuArgGluLeuGlnLeuArg 100
257 TGGAGGAGATGAGAGATCCGGAGAAAGCTTAGGAGCTACAGCTGAGA 306
101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisAspHis 117
307 AATTGTCTACGCATCTTATGGGGAGCTGTCTAACCCACCACGATACCA 356
117 sAspGluPheCysLeuMetPro 124
357 TGATGAATTCCTCCCTTATGCGCT 378

seq_name: gb_est1:AM476468

seq_documentation_block:
LOCUS AM476468 628 bp mRNA EST 24-FEB-2000
DEFINITION ug76d02.y1 NCI-CGAP_Lu33 Mus musculus cDNA clone IMAGE:2937027 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.

ACCESSION AM476468
VERSION AM476468.1 GI:7046574
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 628)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: coapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
pva sequencing by: Washington University Genome Sequencing Center
pva distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium, lhl.gov/bbrp/image/image.htm

Seq primer: -40RP from Gibco
H quality sequence stop: 456.
Location/Qualifiers
1..628
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:2937027"
/clone_lib="NCI-CGAP_Lu33"
/tissue_type="pooled lung tumors"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA was prepared from mRNA obtained from pooled lung tumors with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGGAGCGCGCGAAATTTTTTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

FEATURES
source

141 ATGCCCAATGTCACACGAGAAACGAGAGATGGAGCCCTGCAGAA 190
17 nGlyGluGluAspArgProValGlyGlyGlyGluGlyHisGlnProAlaG 34
191 TGGACAGGAAGACCGCCCTGTGGGAGAGGTGAGGCCACACGCTGCTG 240
34 lyAsnAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisAsnHis 50
241 CAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 290
51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsn 67
291 GGCAGGCTCGCGAGCTTCCCTTAACCTCCGATGGCGCATTCACCA 340
67 gGlnMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheM 84
341 GCAGATGAATGATGCGGTGGTGAGATGGAGATGATATGGAATGTTCA 390
84 eGluGluMetArgGluLeuArgArgLysLeuArgGluLeuGlnLeuArg 100
391 TGGAGGAGATGAGAGATCCGAGAAAGCTTAGGAGCTACAGCTGAGA 440
101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisAspHis 117
441 AATTGTCTACGATCTTATGGGAGCTGTCTAACCAACCAACCAACCA 490
117 sAspGluPheCysLeuMetPro 124
491 TGATGAATTCGCTTATGCCT 512

seq_name: gb_est1:BE334877

seq_documentation_block:
LOCUS BE334877 585 bp mRNA EST 14-JUL-2000
DEFINITION US90b10.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
IMAGE:3325627.5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
KD PROTEIN HGR74 (HUMAN);, mRNA sequence.

ACCESSION BE334877
VERSION BE334877.1 GI:9208653
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 585)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1069791
Seq primer: -4ORP from Gibco
High quality sequence stop: 465.

FEATURES
SOURCE

1..585
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3325627"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 162 a 157 c 172 g 93 t 1 others
ORIGIN

alignment_scores:

Quality: 688.00 Length: 124
Ratio: 5.548 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.387

alignment_block:

US-09-327-750D-12 x BE334877

Align seg 1/1 to: BE334877 from: 1 to: 585

1 MetAlaAsnValHisGlnGluAsnGluGluMetGluGlnProLeuGlnAs 17
166 ATGGCAATGTCCACACGAGAAACGAGAGATGGAGCCCTGCAGAA 215
17 nGlyGluGluAspArgProValGlyGlyGlyGluGlyHisGlnProAlaG 34
216 TGGACAGGAAGACCGCCCTGTGGGAGAGGTGAGGCCACACGCTGCTG 265
34 lyAsnAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisAsnHis 50
266 CAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 315
51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsn 67
316 GGCAGGCTCGCGAGCTTATGCCTTAACCTCCGATGGCGCATTCACCA 365
67 gGlnMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheM 84
366 GCAGATGAATGATGCGGTGGTGAGATGGAGATGATATGGAATGTTCA 415
84 eGluGluMetArgGluLeuArgArgLysLeuArgGluGluGlnLeuArg 100
416 TGGAGGAGATGAGAGATCCGAGAAAGCTTAGGAGCTACAGCTGAGA 465
101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisAspHis 117
466 AATTGTCTACGATCTTATGGGAGCTGTCTAACCAACCAACCAACCA 515
117 sAspGluPheCysLeuMetPro 124
516 TGATGAATTCGCTTATGCCT 537

seq_name: gb_est2:W64711

seq_documentation_block:

LOCUS W64711 590 bp mRNA EST
DEFINITION md71g01.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
clone IMAGE:373872.5' similar to gb:M38188 OVARIAN GRANULOSA CELL
13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.

ACCESSION W64711

VERSION W64711.1 GI:1372353

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 590)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

10-JUN-1996

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:235304

Seq primer: ETPRIMER

High quality sequence stop: 348.

Location/Qualifiers
1. 590

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:373872"

/clone_lib="Soares mouse embryo NbMEL13.5 14.5"

/sex="unknown"

/tissue_type="embryo"

/dev_stage="13.5-14.5dpc total fetus"

/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTGGAGCGCGCGGAATTTTTTTTTTTTTTTTTT

T 3'], on equal amounts of mRNA from 2 13.5dpc and 2

14.5dpc embryos [total RNA provided by Minoru Ko, Wayne

Eco RI adaptors (Pharmacia), digested with Not I and

cloned into the Not I and Eco RI sites of the modified

pT73 vector. Library went through one round of

normalization, and was constructed by Bento Soares and

M. Fatima Bonaldo.

BASE COUNT 165 a 155 c 166 g 104 t

ORIGIN

alignment_scores:

Quality: 688.00 Length: 124
Ratio: 5.548 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.387

alignment_block:

US-09-327-750D-12 x W64711

Align seg 1/1 to: W64711 from: 1 to: 590

```

1 MetAlaAsnValHisGlnGluAsnGluMetGluGlnProLeuGlnAs 17
|||||
134 ATGCCCAATGTCACACAGAAACGAAGAGATGGAGCCCTGCAGAA 183
|||||
17 nGlyGluGluAspArgProValGlyGlyGlyHisGlnProAlaG 34
|||||
184 TGGACAGGAAGACCGCCCTGTGGAGAGGTGTAGGCGCCACCGCTG 233
|||||
34 LyAsnAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisArg 50
|||||
234 CAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAG 283
|||||
51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTTPAlaIleProAsnAr 67
|||||
284 GCCCAGGCTGCCGACTTCCCTTCCGATGGCGCATTCCTCCCAACAG 333
|||||
67 gGlnMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheM 84
|||||
334 GCAGATGAATGACGGTGTGGTGGAGATGAGATGATATGGAATGTTCA 383
|||||
84 eGluGluMetArgGluIleArgArgLysLeuArgGluLeuArg 100
|||||
384 TGGAGAGATGAGAGATGCCGAGAAGCTTAGGAGCTACAGCTGAGA 433
|||||
101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisAspHis 117
|||||
434 AATTGCTACGCATCTTATGGGAGGTGCTTAACCAACCAACCAACCA 483
|||||
117 sAspGluPheCysLeuMetPro 124

```

|||||
484 TGATGAATTCGCTTATGCT 505

seq_name: gb_est2:BG088461

seq_documentation_block:

LOCUS BG088461 599 bp mRNA EST 26-JAN-2001
DEFINITION H3153D07-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone

ACCESSION BG088461

VERSION BG088461

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 599)

AUTHORS Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka

,T.S., Carter,M.G. and Ko,M.S.H.

TITLE Verification and initial annotation of NIA mouse 15K cDNA clone set

JOURNAL Unpublished (2001)

COMMENT Other_ESTs: H3153D07-3

Contact: George J. Kargul

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@igsun.grc.nia.nih.gov

This clone set has been freely distributed to the community. Please

visit http://igsun.grc.nia.nih.gov/cDNA/15k.html for details.

Plate: H3153 row: D column: 07

Seq primer: -21M13 Reverse

High quality sequence stop: 599

POLYA-No.

Location/Qualifiers

1. 599

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="H3153D07"

/clone_lib="NIA Mouse 15K cDNA Clone Set"

/sex="Clones arrayed from a variety of cDNA libraries"

/dev_stage="Clones arrayed from a variety of cDNA

libraries"

/lab_host="DH10B"

/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This

clone is among a rearranged set of 15,247 clones from 11

embryo cDNA libraries (including preimplantation stage

embryos from unfertilized egg to blastocyst, embryonic

part of E7.5 embryos, extraembryonic part of E7.5 embryos

, and E12.5 female mesonephros/gonad) and one newborn

ovary cDNA library. Average insert size 1.5 kb. All

source libraries are cloned unidirectionally with Oligo(dT

)-Not primers. References include: (1) Genome-wide

expression profiling of mid-gestation placenta and embryo

using a 15,000 mouse developmental cDNA microarray. 2000,

Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)

Large-scale cDNA analysis reveals phased gene expression

patterns during preimplantation mouse development, 2000,

Development, 127: 1737-1749; (3) Genome-wide mapping of

unselected transcripts from extraembryonic tissue of

7.5-day mouse embryos reveals enrichment in the t-complex

and under-representation on the X chromosome, 1998, Hum

Mol Genet 7: 1967-1978."

BASE COUNT 163 a 163 c 176 g 97 t

ORIGIN

alignment_scores:

Quality: 688.00 Length: 124
Ratio: 5.548 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.387

alignment_block: